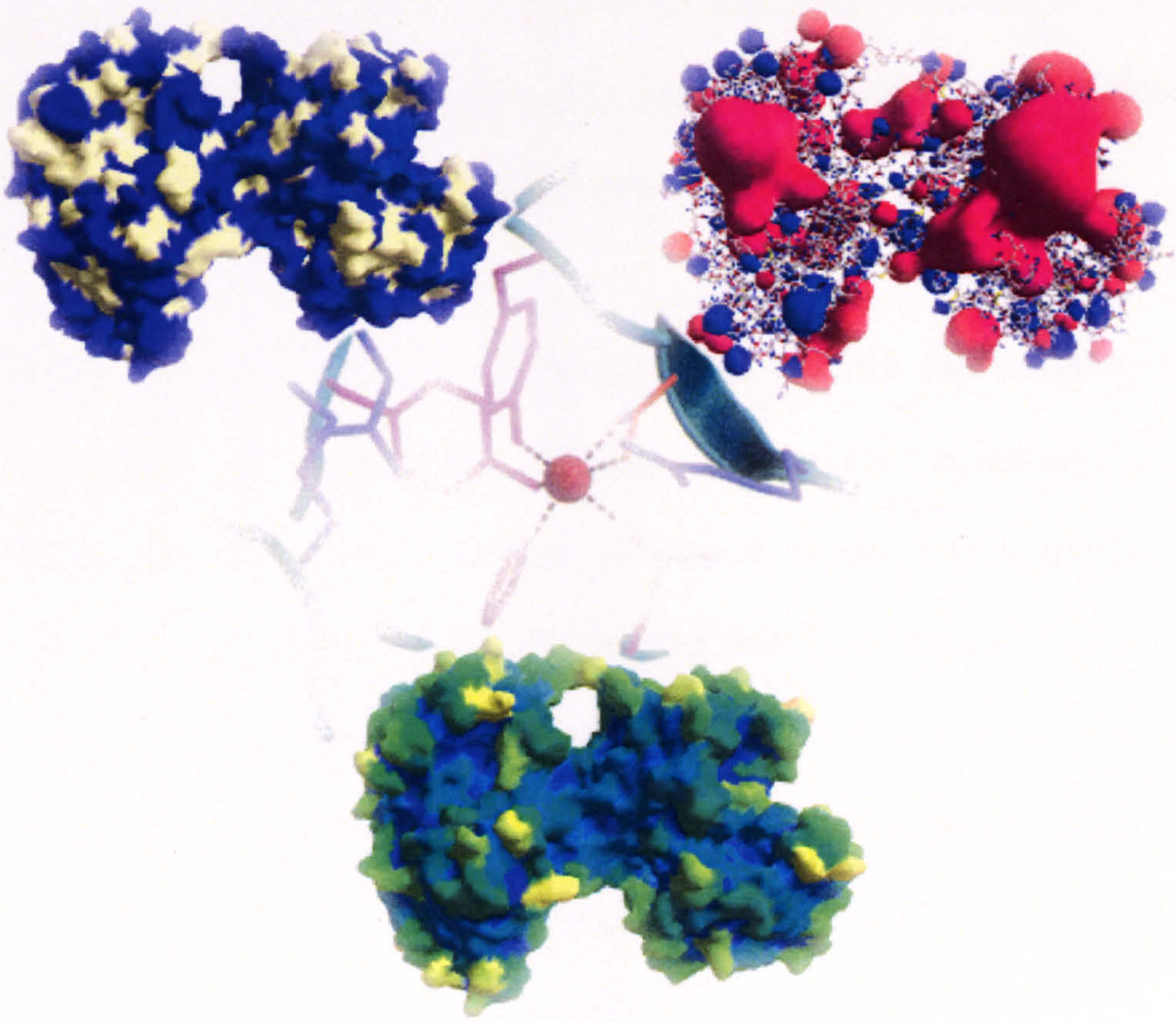
Structure-function studies of native and recombinant transferrins

Sargent, Peter Joseph

Structure-Function Studies of Native and Recombinant Transferrins

by

Peter Joseph Sargent



A thesis submitted for the degree of

Doctor of Philosophy

to the University of London

Metalloprotein Research Group

The Randall Division of Cell and Molecular Biophysics

King's College London, Guy's Hospital, London SE1 1UL

Abstract

The Transferrins are homologous monomeric bilobal glycoproteins of approximately 80kDa that can reversibly exhibit a high affinity for iron. There are four main members of this family: the serum transferrins, involved in iron transport around the body, the lactoferrins that possess a number of functions, including antimicrobial properties, the ovotransferrins and the melanotransferrins. However, the mechanisms by which iron is taken up and released is poorly understood.

The first part of this project was the design of an efficient expression system to obtain routinely the substantive amount of protein necessary for the structure/function study of these proteins, in order to elucidate their mechanism of action and role. The expression of hSTf and hLf was first explored in *Escherichia coli,* where it was shown to be unsuccessful, and then in *Saccharomyces cerevisiae* where, for the first time, a very high level of expression of the full-length hSTf was achieved at levels up to 1.8 mg/ml. The expression of the G394R hSTf mutant was therefore also obtained using the same system.

Following the purification of both recombinant hSTf proteins using a three-step procedure, a four-way structure-function comparison between native WT, native mutant (G394R), recombinant WT and recombinant mutant (G394R) was carried out. Structural identity between these recombinant proteins with their native forms was shown by CD analysis, dynamic light scattering and immunodiffusion assays against anti-hSTf. The ability of the recombinant proteins to bind iron was confirmed by urea-PAGE, UV scanning spectroscopy, along with exhibiting similar

thermodynamic characteristics, calculated through ITC. EPR analysis of the iron-saturated recombinant proteins revealed the characteristic holo-transferrin spectrum.

Biophysical comparisons between the WT and mutant proteins revealed significant changes in the tertiary structure for the variant's C-lobe upon binding iron, which explains a 6-7-fold reduction in the C-lobe affinity for iron. However, such decrease in binding affinity is proposed, with use of homology modelling, not to be due directly to a decreased affinity in the binding site, but more to further interaction of residue R394 elsewhere in the C-lobe. Contradictory to the results published by Evans *et al*, 1994, these interactions are shown not to be involving the residues of the C-lobe iron-binding site, but to be restricting the flexibility of the lobe's two domains therefore preventing them from efficiently closing over the iron atom. Not only has the mutation been shown to affect the C-lobe, but these studies have also shown its subtle repercussions on the N-lobe. It is worth noticing that since the site of these interactions is located near to the regions of the C-lobe involved in interaction with the helical domain of TfR1, it is proposed that they are responsible for the 10-fold reduced affinity of the variant in its diferric form for the TfR1 (Young *et al*, 1984).

Concomitantly to the characterisation, conditions for crystallisation of native hLf and recombinant hSTf were also investigated. Crystals were obtained for both proteins and used in obtaining initial x-ray diffraction data. Such preliminary results, in combination with the experimental characterisation, have brought us undoubtedly one step further towards the full understanding of the mechanism(s) of iron uptake and release of Transferrins.

Dedication

To my Parents.

Thank you for all the love and support you have given me throughout my life and for your constant encouragement during these studies. Without you both, none of this would have been possible and therefore, from the bottom of my heart, I am eternally grateful.

Acknowledgments

I firstly would like to thank Dr Rob Evans and Dr Sebastien Farnaud for giving me the opportunity to work on such an interesting project and also for their continuous encouragement, expertise and time over these last four years. You have both been an enormous inspiration to me.

Secondly, I would like to thank Dr Darrell Sleep, Dr Chris Finnis and all from Delta Biotechnology Ltd for their collaboration and financial contribution, and for giving me the opportunity to experience the biotechnology industry.

I would also like to thank the following people for their expertise and help, in alphabetical order: Dr Andrew Beavil (King's College London) Light Scattering and CD; Dr Bernaud O'Hara (Birkbeck College London) Crystallography; Dr Claire Naylor (Birkbeck College London) Crystallography; Dr. Heinz Zoller (Cambridge Univeristy) Cell Proliferation; Prof. John Ladbury (University College London) ITC; Dr Phil Marsh (King's College London) Molecular Biology; Prof. Richard Cammack (King's College London) EPR; Mr. Sanjay Nilapwar (University College London) ITC and Dr Tam Bui (King's College London) CD Spectroscopy.

Thanks also to Alpesh Patel, John Oakhill and Maryam Amini for providing me with so many fond memories. As well as learning lots from our 'deep' scientific arguments, I had lots of fun working with you.

Finally, I would like to acknowledge the Medical Research Council for funding this project and KCL Enterprises for their support.

Abbreviations

A	Alanine
bp	Base Pair
BSA	Bovine Serum Albumin
C	Cysteine
CD	Circular Dichroism
cv	Column Volume
D	Aspartate
Dcyt b	duodenal cytochrome b
DMSO	Dimethyl sulfoxide
DMT1	Divalent Membrane Transporter 1
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
E	Glutamate
EDTA	Ethylenediaminetetra-acetic acid
EPR	Electron Paramagnetic Resonance
F	Phenylalanine
FITC	Fluorescein Isothiocyanate
FRDA	Friedreich's Ataxia
FPLC	Fast Performance Liquid Chromatography
G	Glycine
H	Histidine
HLf	Human Lactoferrin
HPLC	High Performance Liquid Chromatography
hSTf	Human Serum Transferrin
I	Isoleucine
IgG	Immunoglobulin G
IPTG	Isopropyl-β-D-thiogalactopyranoside
IRE	Iron Responsive Elements
IRP	Iron Responsive Proteins
ITC	Isothermal Titration Calorimetry
K	Kelvin
K	Lysine
Kb	Kilobase
L	Leucine
LB	Luria Bertani Broth
Lbp	Lactoferrin binding protein
Lf	Lactoferrin
M	Methionine
mS	Milli Siemens
MW	Molecular Weight
N	Asparagine
OD	Optical Density
P	Proline
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PEG	Polyethylene glycol
Q	Glutamine
R	Arginine
RNA	Ribonucleic Acid
RT	Room Temperature
rSTf	Rabbit Serum Transferrin
S	Serine
SBA	Sideroblastic anaemia
SD	Standard Deviation
SDS	Sodium dodecyl sulphate
T	Threonine
Tbp	Transferrin binding protein
TEMED	N, N, N', N' - Tetramethylethylenediamine
Tf	Transferrin
TfR	Transferrin Receptor
Tris	Tris-hydroxymethylaminomethane
UTR	Untranslated Regions
UV	Ultraviolet
V	Valine
(v/v)	Volume / Volume
W	Tryptophan
WT	Wild-Type
(w/v)	Weight / Volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
Y	Tyrosine

Websites

Tools

ExPASy Proteomics server	us.expasy.org/
NCBI	www.ncbi.nlm.nih.gov/
NCBI BLASTp	www.ncbi.nlm.nih.gov/BLAST/
ClustalW	www.ebi.ac.uk/clustalw/
APSSP2	www.imtech.res.in/raghava/apssp/

Suppliers

Amersham Biosciences	www1.amershambiosciences.com
BDH Chemicals Ltd	www.bdh.com
BIO101 (Qbiogene)	www.qbiogene.com
BioRad	www.bio-rad.com
Difco Laboratories	www.bd.com
Hampton Research	www.hamptonresearch.com
Millipore / Amicon	www.millipore.com
NCBI	www.ncbi.nlm.nih.gov
New England BioLabs Inc	www.neb.uk.com
Novagen	www.merckbiosciences.co.uk
Promega	www.promega.com
Roche Applied Sciences	www.roche-applied-science.com
Scipac	www.scipac.com
Sigma Aldrich	www.sigmaaldrich.com
Sigma Genosys	www.genosys.co.uk
Stratagene	www.stratagene.com
Tosoh Biosciences	www.tosoh.com
Qiagen	www1.qiagen.com
VWR	uk.vwr.com

Contents

List of Figures

List of Tables

1. Introduction

1.1. Iron

With rare exceptions, virtually all studied organisms from *Archaea* to man are dependent on iron for survival (Aisen *et al,* 2001). One organism found not to require iron is *Lactobacillus plantarum* (Archibald, 1983). These metabolic functions include: provision of a specific binding site for oxygen in the Haem moiety of haemoglobin and conferment of redox activity on the cytochromes of the respiratory chain, and on numerous other enzymes, some of which are involved in DNA synthesis (Andrews, 2000). Iron's ability to confer redox activity is due to it possessing unfilled d atomic orbitals and is able to undergo changes in oxidative states (Fe^{2+}, ferrous and Fe^{3+} ferric) involving one electron. However, this characteristic can also lead to the formation of unstable intermediates with unpaired electrons, i.e. free radical formation. Examples of these include the generation of the superoxide radical O_2^- by the reaction $O_2 + Fe^{2+} \rightarrow O_2^- + Fe^{3+}$ and the formation of $OH\cdot$ (hydroxyl) radical through the Fenton reaction ($Fe^{2+} + H_2O_2 \rightarrow OH\cdot + Fe^{3+} + OH^-$). It is important to note that H_2O_2 in the Fenton reaction is itself generated by superoxide dismutase catalysing $2O_2^- \rightarrow O_2 + O_2^{2-}$, then O_2^{2-} reacting with $2H^+$. These hydroxyl radicals are so reactive they will attack and destroy cell membranes, as well as DNA. This shows that iron homeostasis needs to be tightly controlled, as even though iron is essential for metabolism, it is also toxic in its free form.

1.2. Iron Homeostasis

As previously discussed, iron is an important requirement in our diet. Cellular iron deficiency is known to arrest cell growth and ultimately lead to cell death (Hentze *et al,* 2004). Because of its importance, organisms have evolved highly sophisticated mechanisms to control iron homeostasis in order to provide the amounts required for life, but at the same time, minimising its toxic effect. In mammals, this regulation is controlled at many levels.

1.2.1. Iron Requirement

The daily iron requirement for a healthy individual is approximately 25mg. This varies from person to person. For instance, growing children and menstruating women have much higher iron demands. This demand increases even more for pregnant individuals. Iron requirements are three times higher in pregnancy than they are in menstruating women (Tapiero *et al,* 2001). On average though, the production of 200 billion new erythrocytes every day requires ~20mg of iron for haemoglobin synthesis, accounting for nearly 80% of the iron demand (Hentze *et al,* 2004). The remaining 20% of the iron demand is used as a constituent in the production of other haemoproteins, iron-sulphur (Fe-S) proteins, and proteins that use iron in other functional groups.

A typical Western diet provides on average 6mg of this iron (both haem and non-haem) per 4120 kJ of energy intake (Beard *et al,* 2000). However, the amount of iron absorbed is regulated depending on the iron status of the individual. An individual with high levels of iron will not absorb iron so readily as someone with low levels. Also, the form of iron is important. Food iron occurs in two forms, haem iron and

non-haem iron. Nearly 50% of iron in meat is of the haem form (Tapiero *et al,* 2001). This form is the most important dietary source of iron, as it is more readily absorbed than non-haem iron. This is why vegetarians can be at a relatively greater risk of iron deficiency (Beard *et al,* 2000).

The majority of this iron is not utilised straight away, but accumulates safely in the body. To give an idea how much iron does accumulate, iron represents approximately 35 and 45mg/kg of body weight in adult women and men, respectively (Leiu *et al,* 2001). The majority of the total body iron, 60-70%, is present in haemoglobin in circulating erythrocytes, 10% in the form of myoglobin, cytochromes, and iron-containing enzymes, and the remaining 20-30% is stored in ferritins and haemosiderins in hepatocytes and reticuloendothelial macrophages (Conrad et al, 1999). This distribution of iron does vary. For example, in situations of iron deficiency, erythroid precursors have priority and the production of red blood cells continues at the expense of other tissues.

The body is economical in its handling of iron. Only 0.5-2mg enters and leaves the body on a daily basis (Andrews, 2000). The majority of the iron is recycled. For instance, when red blood cells reach the end of their lifespan, they are phagocytosed by reticuloendothelial macrophages, which recycle the iron from haem in order to be used in the production of more erythrocytes, or to be used else where in the body.

1.2.2. Iron Absorption and Transport

Intestinal absorption of iron is thought to be the critical step in iron homeostasis (De Freitas *et al,* 2001). The enterocytes found in the epithelial cell layer of the duodenum are specialised for absorption and transportation of iron. These cells differ from other cells in that they express a number of proteins related to iron uptake and transport (Figure 1.1). This figure shows a model for iron transport across duodenal epithelial cell proposed by Andrews, 2000.

It is proposed that non-haem iron (Fe^{3+}) in the duodenum is reduced by a ferric reductase (Dcytb: duodenal cytochrome b) in the brush border, before being transported into the enterocyte through the transmembrane iron transporter DMT1 (divalent metal transporter 1). Once inside the cell, the iron (Fe^{2+}) is either stored in ferritin, or passes through the basolateral membrane to reach the plasma by an iron exporter known as ferroportin. This step is thought to be followed by the reoxidation of Fe^{2+} by the plasma ferroxidase ceruloplasmin (or its membrane-associated homologue, hephaestin) (Pantopoulos, 2004). Fe^{3+} is then picked up by apo-transferrin, to make holo-transferrin, through a mechanism that is not yet understood. In normal individuals, this extracellular iron circulates in the plasma bound to serum transferrin. Iron, bound in this way is no longer toxic, in that it no longer has the ability to produce free radicals.

1.2.3. Iron Uptake

In vertebrates, the major protein involved in iron uptake of iron-loaded transferrin into cells is the plasma membrane transferrin receptor (TfR) (De Freitas *et al,* 2001). Two types of TfR are known, TfR1 and TfR2. The latter molecule has only recently

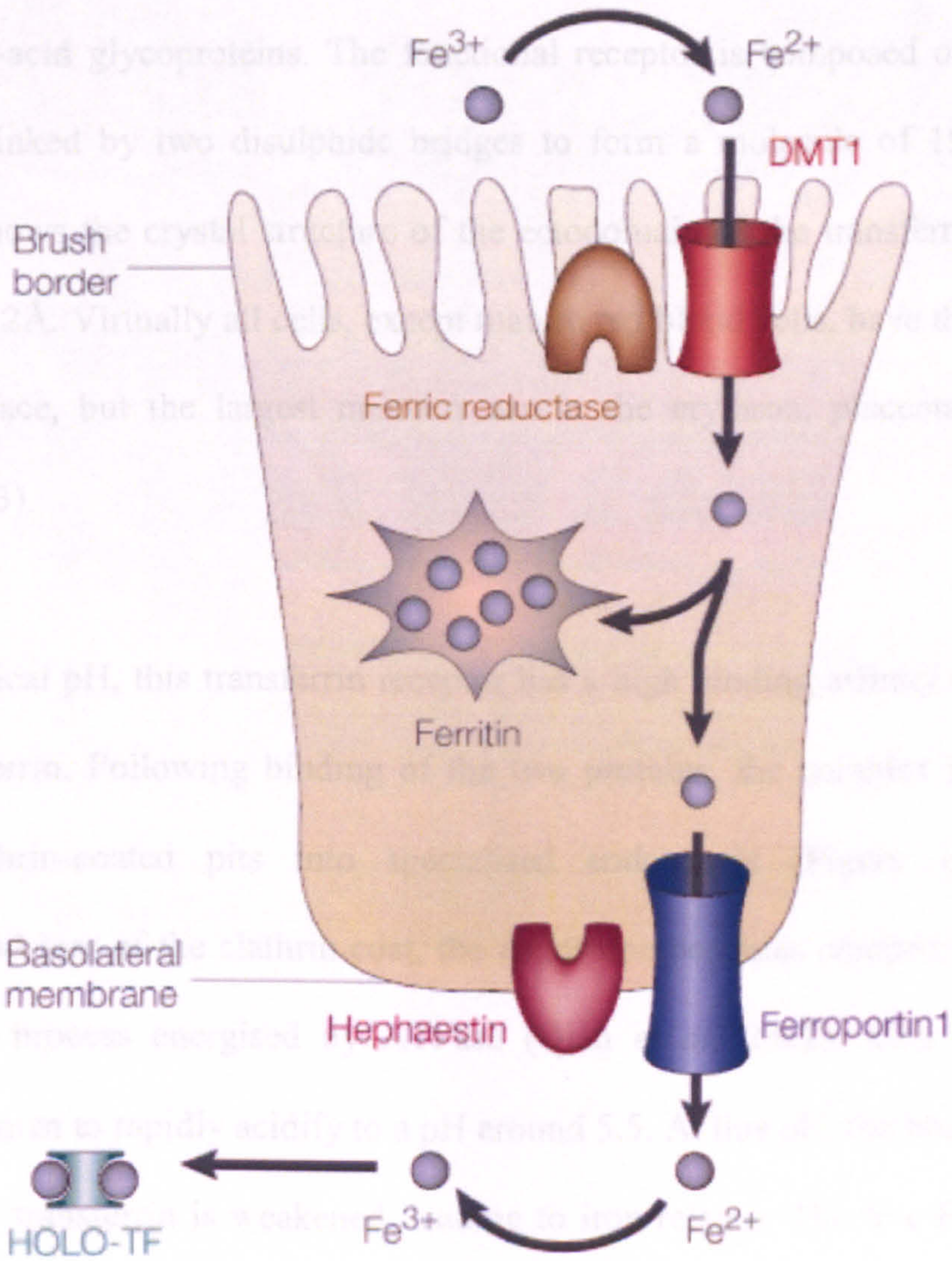


Figure 1.1. A model for iron transport across duodenal epithelial cells, proposed by Andrews, 2000. Shown are the proteins thought to be involved in the absorption and transport of iron from the diet into the body's circulatory system.

been identified and its function remains unclear (Kawabata *et al,* 1999). What is known, is that both receptors bind diferric transferrin in a pH-dependent manner and delivers iron to the cells. However, TfR2 has a much lower binding affinity (25 times lower) and its distribution around the body varies significantly (Trinder and Baker, 2003). The best characterised of the two is TfR1 (Cheng et al, 2004). This receptor is a 760-amino-acid glycoproteins. The functional receptor is composed of two such monomers, linked by two disulphide bridges to form a molecule of 190,000 Da. Figure 1.2 shows the crystal structure of the ectodomain of the transferrin receptor resolved at 3.2Å. Virtually all cells, except mature red blood cells, have this receptor on their surface, but the largest numbers are in the erythron, placenta and liver (Beguin, 2003).

At physiological pH, this transferrin receptor has a high binding affinity for diferric serum transferrin. Following binding of the two proteins, the complex internalises through clathrin-coated pits into specialised endosomes (Figure 1.3). Upon maturation and loss of the clathrin coat, the endosome becomes competent to pump protons in a process energised by ATPase (Qian *et al,* 2002). This causes the endosomal lumen to rapidly acidify to a pH around 5.5. At this pH, the binding of the iron to serum transferrin is weakened, leading to iron release. The free Fe^{3+} is then reduced to Fe^{2+} on the *cis*-side of the endosomal membrane, before being transported out of the endosome by DMT1 (section 1.2.2). Once in the cytosol, iron is either stored or utilised by the cell. Following release of the iron, the transferrin-TfR complex is then recruited through exocytic vesicles back to the cell surface. At physiological pH, apo-transferrin is dissociated from the receptor and enters back

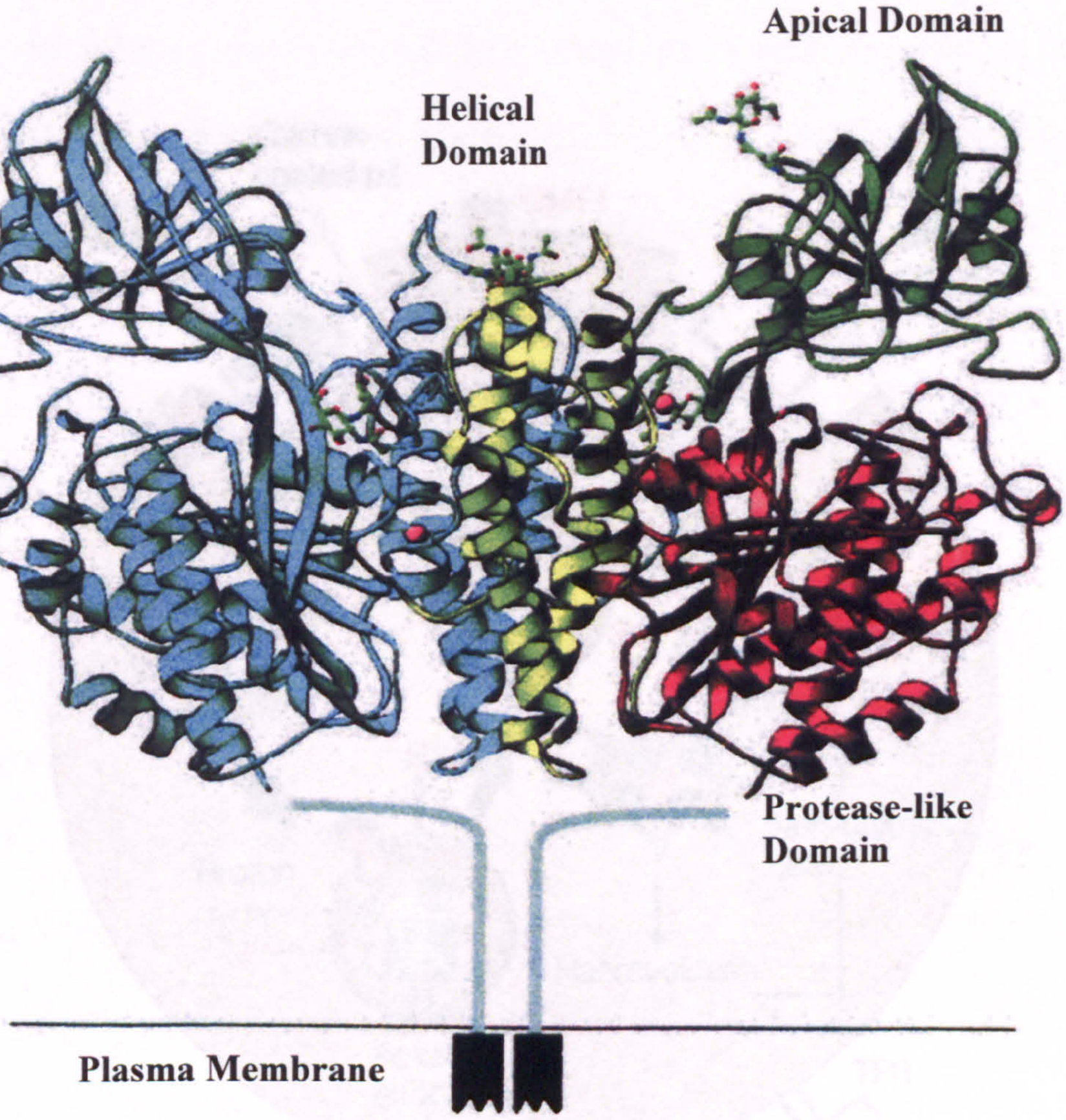


Figure 1.2. The crystal structure of the ectodomain of TfR resolved at 3.2 Å. *(Lawrence et al, 1999)*. Each TfR monomer can be divided up into a number of domains; Cytoplasmic domain, N-terminus consisting of 61 amino acids; Transmembrane domain, 26 amino acids long hydrophobic region; and the transferrin binding domain, 671 amino acids long. One monomer is depicted in blue, whilst the other monomer is divided up into the protease domain, apical domain and helical domain, shown in red, green and yellow, respectively.

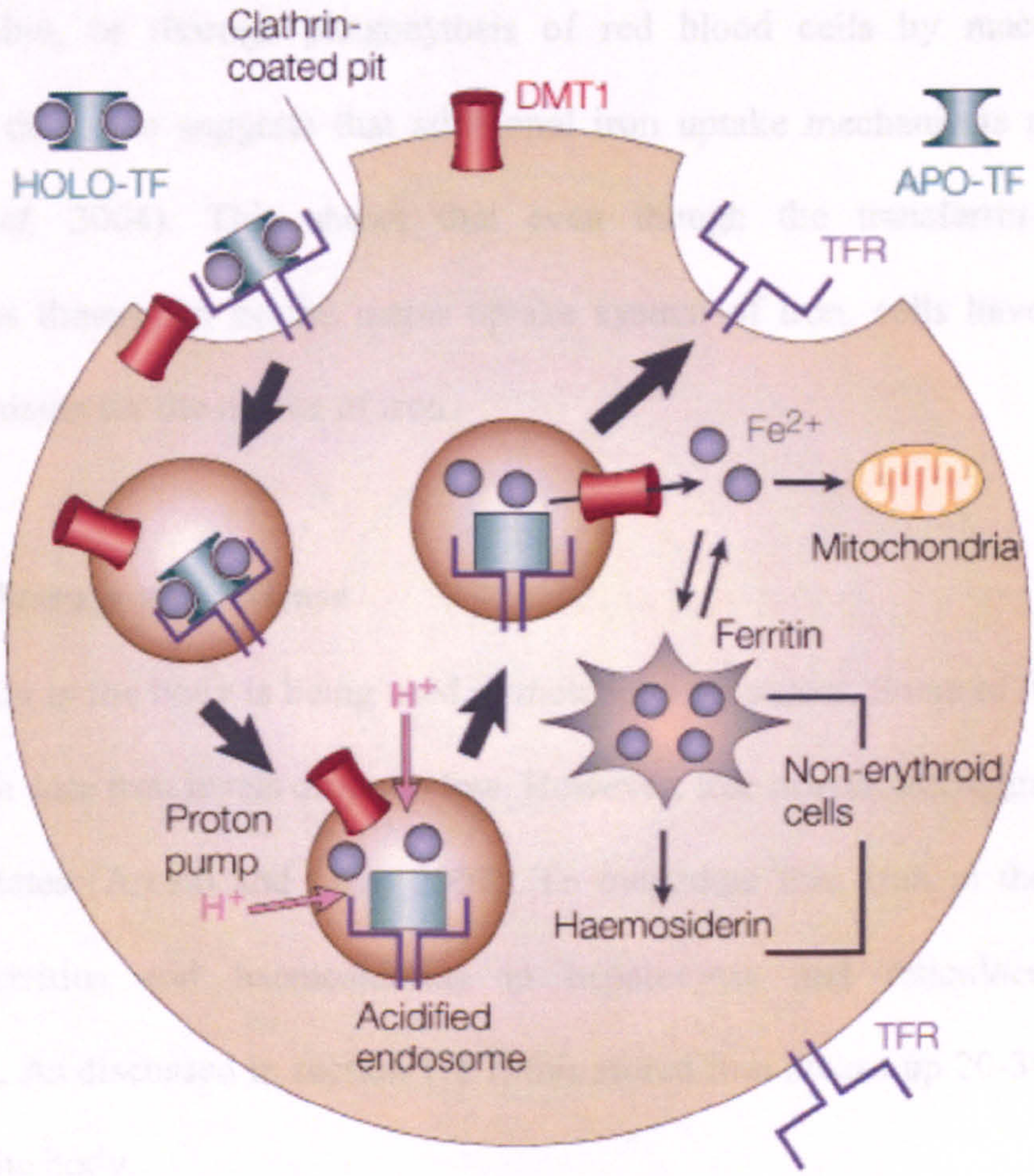


Figure 1.3. A model for iron uptake via the transferrin receptor-mediated endocytosis pathway *(Andrews, 2000).* Shows the internalisation of diferric transferrin, following binding to TfR1, followed by the release of iron from transferrin and its transport into the cytosol for either storage or utilisation. Finally, the transport of the transferrin-TfR1 complex to the cell membrane and the subsequent release of transferrin back into circulation.

into the circulation, and reutilised (Morgan, 1996). Transferrin has been calculated to be recycled between 100-200 times during its lifetime (Aisen *et al,* 2001).

Cells may also take up iron in other protein-bound forms, such as acidic isoferritin, or haemoglobin, or through phagocytosis of red blood cells by macrophages. Biochemical data also suggests that additional iron uptake mechanisms may exist (Hentze *et al,* 2004). This shows that even though the transferrin-mediated endocytosis is thought to be the major uptake system of iron, cells have evolved many mechanisms for the uptake of iron.

1.2.4. Iron Storage and Release

Not all the iron in the body is being used in metabolic processes. Some of it is stored as a reserve in case iron levels drop too low. However, free iron would aggregate into toxic precipitates (Arosio and Levi, 2002). To overcome this, iron in the body is stored in ferritins and haemosiderins in hepatocytes and reticuloendothelial macrophages. As discussed in section 1.2.1, this stored iron makes up 20-30% of the total iron in the body.

Ferritin is a ubiquitous and highly conserved iron-binding protein. A number of different forms exist. Examples of these include, cytosol ferritin, mitochondrial ferritin and serum ferritin. The structure of the cytosolic form of L-chain horse apoferritin is shown in Figure 1.4. Mammalian ferritins are all heteropolymers of 24 subunits of two types, designated H for heavy (21,000 Da), or heart (but also kidney), the tissue in which it abounds, and L for light (19,500 Da) or liver (but also spleen) (Aisen *et al,* 2001). These molecules have the ability to sequester up to

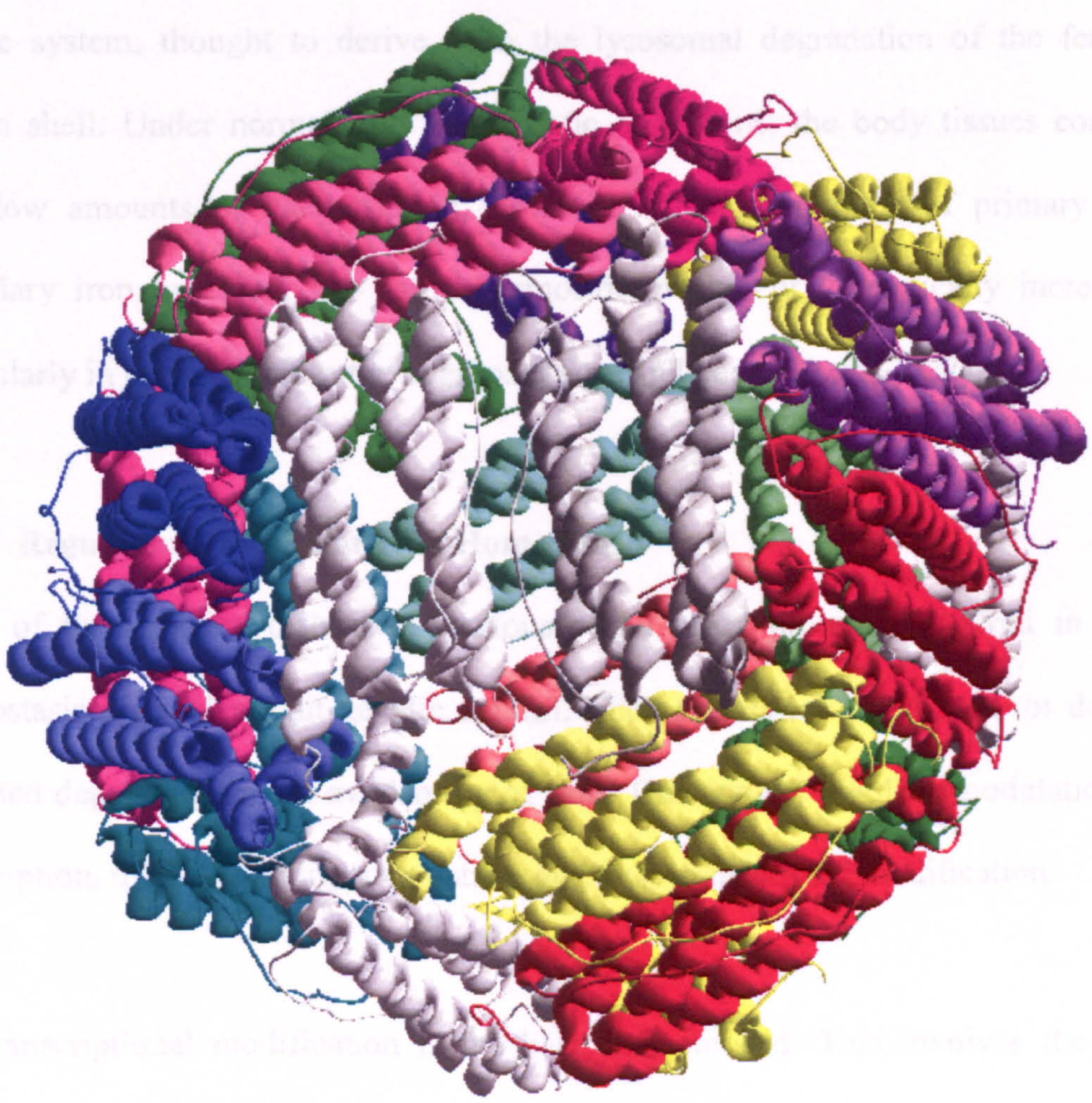


Figure 1.4. A ribbon diagram of L-chain horse apoferritin (1.95Å) (1AEW) created using Deep Viewer / Swiss PDB Viewer. Each of the 24 subunits is coloured differently to give an idea of the complexity of the molecule.

approximately 4500 iron atoms in the form of Fe^{3+} (Chasteen and Harrison, 1999). An important feature of ferritin is its ability to catalyse the oxidation of Fe^{2+}, the predominant iron form in the cytosol, converting it to Fe^{3+} before it is sequestered in the ferritin mineral core.

Haemosiderin is a poorly defined iron-protein complex that forms an insoluble iron storage system, thought to derive from the lysosomal degradation of the ferritin protein shell. Under normal iron homeostatic conditions, the body tissues contain only low amounts of haemosiderin; however, under conditions of primary and secondary iron overload, the tissue haemosiderin content dramatically increases, particularly in organs such as the liver, pancreas, and heart (Testa, 2002).

1.2.5. Regulation of Cellular Iron Homeostasis

Many of the proteins discussed previously and other proteins involved in iron homeostasis are under tight genetic control. Their expression can be up- or down-regulated depending on the availability of iron. This control involves modulation of transcription, mRNA stability, translation, and post-translational modification.

Post-transcriptional modification is the best characterised. This involves the iron regulatory proteins (IRPs) and iron responsive elements (IREs). IRP1 and IRP2 are involved in the coordinated post-transcriptional regulation of iron metabolism by binding to IREs (Wang *et al,* 2002). These IREs are stem-shaped structures located in the 5' and 3' untranslated regions (UTRs) of mRNA. Their structure comprises a terminal hexanucleotide loop region with the sequence 5'-CAGUGX-3', where X can be A, C, or U, but never G, and a base-paired stem structure that is interrupted by

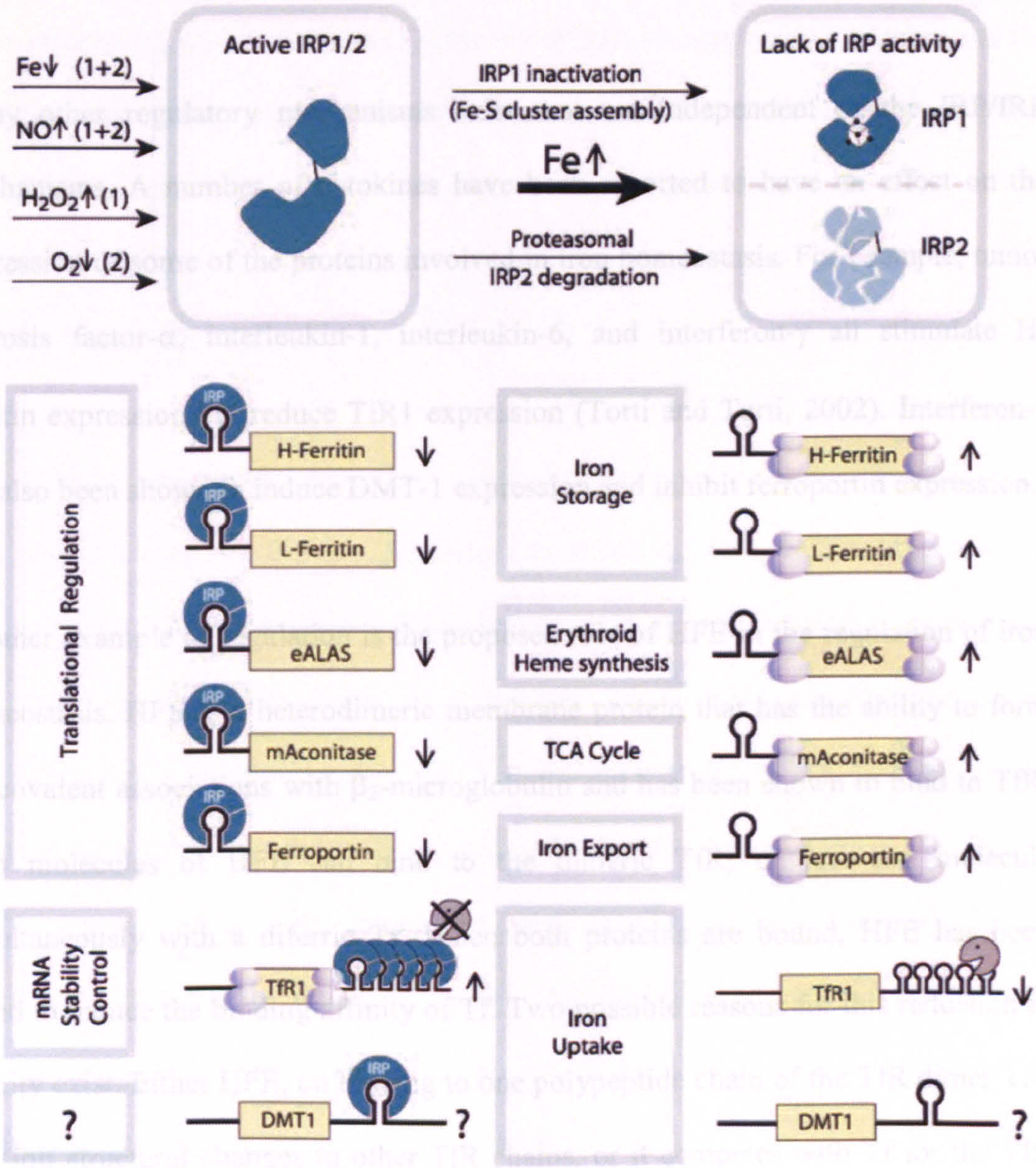
a conserved unpaired cytosine, six nucleotides down from the loop. Table 1.1 shows examples of some of the proteins under IRE control. These include proteins involved in iron storage, erythroid heme synthesis, the TCA cycle, iron export and iron uptake.

The binding of IRPs to single IREs in the 5' UTRs of mRNA has been found to block translation. Whereas, binding of IRPs to IREs in the 3' UTRs stabilises the mRNA, therefore having a positive effect on translation. Figure 1.5 shows some of the proteins regulated in such a way, and also the effect iron levels have on their translation.

The actual binding of IRPs to IREs is regulated by a number of different factors. When iron levels are high, a cubane iron sulphur cluster assembles in IRP1, inhibiting IRE binding and converting IRP1 to an aconitase (Pantopoulos, 2004). Under iron limiting conditions this iron sulphur cluster does not form, and IRP1 is able to bind to IREs. This is different for IRP2 as this protein does not contain an iron sulphur cluster. IRP2 has been found to contain an iron-dependent degradation domain. In the presence of iron, proteosomes bind to this protein and degrade it, thus preventing its interaction with IREs. Even though these two IRPs have been identified, little is known about their individual contribution to the maintenance of cellular iron homeostasis. In *in vitro* experiments, it has been shown that both proteins do not differ in their ability to be able to regulate ferritin and TfR expression (Erlitzki *et al,* 2002). However, Meyron-holtz *et al*, 2004, suggest that IRP2 dominates post-transcriptional regulation of iron metabolism in mammals. The actual control of IRP expression is poorly understood. A number of factors that influence their expression are known. These include the influence of reactive oxygen species,

Gene	IRE location	Function	Reference
Single IREs			
Ferritin L chain	5'	Iron Store	Hentze and Kühn, 1996
Ferritin H chain	5'	Iron Store	
Mitochondrial Aconitase	5'	Enzyme in Citric Acid cycle	
Erythroid 5-ALA-synthase	5'	Specific key enzyme in heme synthesis	
Nramp2	3'	Iron Importer	Gunshin *et al*, 1997
Ferroportin 1 (IREG1)	5'	Iron Exporter	Mok *et al*., 2004
DMT1	3'	Iron Transporter	Hentze, 2004
Succinate dehydrogenase	5'	Involved in the citric acid cycle	Pantopoulos, 2004
Multiple IREs			
TfR1	3'	Trf-binding and – transport	Hentze and Kühn, 1996

Table 1.1. Examples of Iron Responsive Element (IRE)-containing genes and their function. The location of the IRE, whether in the 3' or 5' untranslated region (UTR), is also stated. Only the TfR1 mRNA contains multiple IREs.



Hentze *et al*, 2004

Figure 1.5. The IRP-IRE regulation system showing the effect iron, NO, H_20_2 and O_2 levels has on the binding of IRPs to IREs, and the subsequent effect this has on translation on a number of proteins. IRPs binding to IREs in the 3' UTR portion of mRNA has been found to protect the mRNA from degradation, whereas, the binding of IRPs to IREs in the 5' UTR portion has been found to block translation.

shown to disassemble the iron sulphur cluster of IRP1, as well as nitric oxide and hypoxia having an effect on both IRPs (Hentze *et al*, 2004).

Many other regulatory mechanisms exist that are independent of the IRP/IRE mechanisms. A number of cytokines have been reported to have an effect on the expression of some of the proteins involved in iron homeostasis. For example, tumor necrosis factor-α, interleukin-1, interleukin-6, and interferon-γ all stimulate H-ferritin expression but reduce TfR1 expression (Torti and Torti, 2002). Interferon-γ has also been shown to induce DMT-1 expression and inhibit ferroportin expression.

Another example of regulation is the proposed role of HFE in the regulation of iron homeostasis. HFE is a heterodimeric membrane protein that has the ability to form noncovalent associations with β_2-microglobulin and has been shown to bind to TfR. Two molecules of HFE can bind to the dimeric TfR, or just one molecule simultaneously with a diferric Tf. When both proteins are bound, HFE has been found to reduce the binding affinity of Tf. Two possible reasons for this reduction in affinity exist. Either HFE, on binding to one polypeptide chain of the TfR dimer, can transmit structural changes to other TfR chains, or it competes with Tf for the TfR binding site (Giannetti *et al,* 2004*)*. However, how HFE actually contributes to the regulation of cellular iron homeostasis is still unknown. Roy *et al*, 1999 showed that the expression of HFE in the cell reduces iron uptake by 33%, but does not affect the endocytic or exocytic rates of TfR cycling.

1.2.6. Iron Metabolism Disorders in Humans

Iron metabolism-related diseases in humans fall into three major categories: (I) diseases associated with defective regulation of iron absorption, (II) diseases caused by erroneous tissue and/or subcellular compartmentalisation of iron, and (III) secondary disorders induced by altered iron content in cells and/ or other tissues (Lieu *et al*, 2001). Just a few examples of these diseases are listed in Table 1.2, along with the gene involved and the defect.

Haemochromatosis is the most common genetic disorder known in Caucasians of European ancestry. In the United States, as much as 10% of the population is heterozygous for this condition, and the homozygous state is believed to be as much as 0.5% of the population (Sheth and Brittenham, 2000). In this disorder, the control of iron uptake is abnormal causing an increase in iron uptake and the progressive build-up of body iron, ultimately becoming fatal if not treated. The iron overload firstly accumulated in hepatocytes, but subsequently the iron builds up in the pancreas, heart and other organs. Five known types of haemochromatosis exist. Type 1, or hereditary haemochromatosis, is caused by a mutation in the gene encoding HFE. Even though the exact mechanism is still a mystery, this protein has a unique perinuclear localisation in cells in the deep crypts of the small intestine where it is thought to interact with β_2-microglobulin and aid in iron absorption. HFE has also been found to interact with the TfR. Type II, or Juvenile haemochromatosis, is an autosomal recessive disease very distinct from hereditary haemochromatosis. In this disorder, the onset of iron overload is much earlier and is more severe, with endocrine dysfunction, joint disease and cardiac abnormalities. Even though the exact cause of this disease is unknown, mutations in the hepcidin and haemojuvilin

Disease	Gene	Defect
Aceruloplasminaemia	Ceruloplasmin	Point mutation, insertion, deletion
African dietary overload	-	-
Congenital atransferrinaemia	Transferrin	-
Congenital dyserythropoietic anaemia	-	-
Erythropoietic protoporphyria	Ferrochelatase	Point mutation
Familial hypoferremic microcytic anaemia	-	-
Friedreich's ataxia	FRDA	Point mutation
Hereditary haemochromatosis	Hfe	Point mutation
HFE 3 type haemochromatosis	TfR2	Point mutation
Hyperferritinaemia with cataract	Ferritin L	Point mutation in IRE
Juvenile haemochromatosis	-	-
Menkes disease	ATPase 7A	Point mutation
Porphyria cutanea tarda	URO-D	Point mutation, insertion, deletion
Sideroblastic anaemia	ALAS2	Point mutation
Sideroblastic anaemia and ataxia	ABC7	Point mutation
Wilson's disease	ATPase 7B	Point mutation

Table 1.2. Iron metabolism disorders in humans, the gene involved and the defect. Where the gene and / or the defect is unknown a (–) is used. (Adapted from Lieu *et al*, 2001).

genes are thought to play a role. Type III is a rare form of the disease thought to be cause by mutations in the TfR2 gene and type IV is again rare, and is thought to occur due to mutations in the ferroportin gene. The fifth type is neonatal haemochromatosis. This again is very rare and the exact cause is unknown.

Atransferrinaemia is a rare autosomal recessive disorder in which plasma transferrin is absent, or almost absent in the case of hypotransferrinaemia. Only about eight cases have been described in six families (Beutler *et al*, 2000). A severe microcytic anaemia, present from birth, only responds to transferrin infusion. Without this, or blood transfusions, patients are at great risk of iron-overload and usually die early in life.

Aceruloplasminaemia, like atransferrinaemia, is a newly recognised, rare autosomal recessive disorder. This disorder results from the deficiency, or complete lack of ceruloplasmin ferroxidase activity as a consequence of mutations in the ceruloplasmin gene. Clinically, the disease consists of the triad of adult-onset neurological disease, retinal degeneration and diabetes mellitus (Miyajima, 2003). Ceruloplasmin deficiency is also linked to Wilson's disease and Menke's disease.

Hyperferritinaemia with autosomal congenital cataract is a disorder characterised by early-onset, bilateral nuclear cataracts and moderately elevated serum ferritin concentrations. In contrast with normal individuals, serum ferritin in patients with this disorder is mainly composed of L-ferritin, with H-ferritin levels being sometimes

too low to detect. This disorder is caused by multiple point mutations in the IRE of the L-ferritin mRNA, therefore affecting the regulation of its synthesis.

Friedreich's Ataxia (FRDA) and sideroblastic anaemia (SBA) are disorders that are associated with aberrant mitochondrial iron homeostasis (Roy and Andrews, 2001). In FRDA, patients show evidence of mitochondrial iron overload and loss of activity of iron-sulphur cluster-containing enzymes. They frequently die from cardiomyopathy. This disorder is thought to be due to point-mutations in the gene encoding frataxin, a protein involved in the mediation of mitochondrial iron homeostasis. SBA is quite different. Even though there are many different causes of SBA, they all affect the efficiency of haem production within erythroblast mitochondria. There are two forms of SBA: SBA with ataxia (muscular problems), caused by point mutations in the *ABC7* gene and SBA without ataxia, caused by a mutation in *ALAS2*.

Some disorders actually have a secondary effect on iron metabolism. One such disorder is thalassaemia. This is a disease caused by insufficient synthesis of structurally normal α or β chains of haemoglobin and results in anaemia. There are many different types of thalassaemia and even more genetic mutations that cause it. The two main types are: (I) α-thalassaemia, caused by relative or absolute deficiency of α-chains and (II) β-thalassaemia, caused by relative or absolute deficiency of β-chains. These two groups can be split further into thalassaemia minor, heterozygous α- or β-thalassaemia where anaemia may be present, but is very mild and thalassaemia major, homozygous α- or β thalassaemia causing severe anaemia. Because thalassaemia causes anaemia, patients with this disease, especially in the

major cases, require blood transfusions at frequent intervals throughout their lives to provide them with the haemoglobin they require. This is an effective treatment, but unfortunately the increased uptake of iron, not through the usual regulatory pathways of the gut, but in the form of haemoglobin, leads to iron overload. This is made worse due to the fact that iron does not get recycled for use in synthesis of haemoglobin as in normal patients.

1.3. The Transferrins

1.3.1. Introduction

The transferrins are a family of homologous monomeric bilobal glycoproteins of molecular weight approximately 80kDa, that have an important role in iron homeostasis. Even though the first isolation of a member of the transferrin family (ovotransferrin) came in 1900 (Osborne and Campbell, 1900), it was not until 1944 that it function was elucidated (Schade and Caroline, 1944). Soon after, Schade *et al*, 1946, isolated a similar protein from blood plasma, known now as serum transferrin. This family has now been found to comprise of four main members. The four main members include; the serum transferrins, which are known to transport iron from its site of absorption to the site of utilisation, or storage; the lactoferrins, which have been found to possess antimicrobial properties; the ovotransferrins, which are known to be a major component of egg white; and the melanotransferrins, found on the surface of melanoma cells. Each of these members exhibits a high affinity for iron(III) ($K_d \sim 10^{-22}$M)(Aisen *et al*, 1978) and have been found in a wide range of organisms, in both vertebrate and invertebrate groups. Table 1.3 lists a few of the transferrins that have either been sequenced, or partially sequenced.

	Chordata			Anthropoda			Echinod-ermata
Transferrin	Mammalia	Aves	Amphibia	Actinopterygii	Insecta	Malacostraca	Echinoidea
Serum Transferrin	Human		African clawed toad	Medaka	Mosquito (Aedes)		
	Mouse			Atlantic cod	Cockroach		
	Rat			Haddock	Silkworm		
	Rabbit			Coho salmon	Fruit fly		
	Cow			Cherry salmon	Picture wing fly		
	Horse			Rainbow trout	Tobacco hornworm		
	Pig			Sockeye salmon	Bean bug		
	Possum			Chinook salmon	Flesh fly		
	Guinea Pig			Atlantic salmon	Honeybee		
	Golden Hamster			Halibut	Termite		
	Hippo			Brown trout	Wax moth		
	Bowhead Whale			Brook trout	Tse tse fly		
	Cuvier's Whale			Lake trout			
	Peruvian Whale			Amago			
	Sperm Whale			Japanese char			
	Boutu			Gila trout			
	Harbor porpoise			Common carp			
	Sheep			Goldfish			
				Zebra fish			
Lactoferrin	Human						
	Mouse						
	Cow						
	Horse						
	Pig						
	Water buffalo						
	Camel						
	Goat						
	Rat						
Ovo-transferrin		Duck					
		Chicken					
Melano-transferrin	Human						
	Mouse						
	Chicken						
	Rabbit						
				Rainbow trout			
				Fugu			
Others							Red urchin
							Green urchin
							Sea urchin
							Hawaiin urchin
			Saxiphilin (bullfrog)				
						Pacifastin (crayfish)	

Table 1.3. Transferrins that have been identified and sequenced. Partial and variants are shown in italics. This data was last updated in 2003 from: www.chatham.edu/undergraduate/bio/lambert/transferrin/transferrins.htm

Although the transferrins have evolved to bind iron(III) with high affinity at neutral pH, their metal-binding sites are able to bind a wide range of different metal ions, especially tripositive ones. By 1992 thirty elements, in addition to iron, were known to be able to bind to one or both of the metal-binding sites in hSTf *in vitro* (Welch, 1992). Since then a further four metal ions, cerium(IV) (Smith *et al*, 1994), bismuth(III) (Li *et al*, 1996), lutetium(III)(Harris *et al*, 1999) and ytterbium(III)(Du *et al*, 2002) have been shown to bind to members of the transferrin family of proteins. The elements now proven to bind to transferrins are highlighted in Figure 1.6. As serum transferrin in normal individuals is only partially iron-saturated, the protein, in theory could be involved in the transport of many other metals around the body.

1.3.2. General Structure

The transferrin polypeptide is approximately 670 amino acids long with a molecular weight of 79,550Da (MacGillivray *et al*, 1982) folded into two lobes that exhibit high levels of homology, the N- and C-lobes. The two lobes of human serum transferrin are 48% homologous (Baldwin, 1993) and for both porcine and rabbit serum transferrin, 45.5% and 42.9% respectively (Hall *et al*, 2002). This high degree of homology is consistent with the theory that both lobes arose from a gene duplication process (Williams, 1982).

Within the secondary structure of these proteins, there are six conserved disulphide bridges in each lobe, with an extra three semiconserved bridges in the C-lobe (Williams, 1982). Chicken and duck ovotransferrin have only fifteen bridges,

1 **H** Hydrogen 1.00794																	2 **He** Helium 4.003
3 **Li** Lithium 6.941	4 **Be** Beryllium 9.012182											5 **B** Boron 10.811	6 **C** Carbon 12.0107	7 **N** Nitrogen 14.00674	8 **O** Oxygen 15.9994	9 **F** Fluorine 18.9984032	10 **Ne** Neon 20.1797
11 **Na** Sodium 22.989770	12 **Mg** Magnesium 24.3050											13 **Al** Aluminum 26.981538	14 **Si** Silicon 28.0855	15 **P** Phosphorus 30.973761	16 **S** Sulfur 32.066	17 **Cl** Chlorine 35.4527	18 **Ar** Argon 39.948
19 **K** Potassium 39.0983	20 **Ca** Calcium 40.078	21 **Sc** Scandium 44.955910	22 **Ti** Titanium 47.867	23 **V** Vanadium 50.9415	24 **Cr** Chromium 51.9961	25 **Mn** Manganese 54.938049	26 **Fe** Iron 55.845	27 **Co** Cobalt 58.933200	28 **Ni** Nickel 58.6934	29 **Cu** Copper 63.546	30 **Zn** Zinc 65.39	31 **Ga** Gallium 69.723	32 **Ge** Germanium 72.61	33 **As** Arsenic 74.92160	34 **Se** Selenium 78.96	35 **Br** Bromine 79.904	36 **Kr** Krypton 83.80
37 **Rb** Rubidium 85.4678	38 **Sr** Strontium 87.62	39 **Y** Yttrium 88.90585	40 **Zr** Zirconium 91.224	41 **Nb** Niobium 92.90638	42 **Mo** Molybdenum 95.94	43 **Tc** Technetium (98)	44 **Ru** Ruthenium 101.07	45 **Rh** Rhodium 102.90550	46 **Pd** Palladium 106.42	47 **Ag** Silver 107.8682	48 **Cd** Cadmium 112.411	49 **In** Indium 114.818	50 **Sn** Tin 118.710	51 **Sb** Antimony 121.760	52 **Te** Tellurium 127.60	53 **I** Iodine 126.90447	54 **Xe** Xenon 131.29
55 **Cs** Cesium 132.90545	56 **Ba** Barium 137.327	57 **La** Lanthanum 138.9055	72 **Hf** Hafnium 178.49	73 **Ta** Tantalum 180.9479	74 **W** Tungsten 183.84	75 **Re** Rhenium 186.207	76 **Os** Osmium 190.23	77 **Ir** Iridium 192.217	78 **Pt** Platinum 195.078	79 **Au** Gold 196.96655	80 **Hg** Mercury 200.59	81 **Tl** Thallium 204.3833	82 **Pb** Lead 207.2	83 **Bi** Bismuth 208.98038	84 **Po** Polonium (209)	85 **At** Astatine (210)	86 **Rn** Radon (222)
87 **Fr** Francium (223)	88 **Ra** Radium (226)	89 **Ac** Actinium (227)	104 **Rf** Rutherfordium (261)	105 **Db** Dubnium (262)	106 **Sg** Seaborgium (263)	107 **Bh** Bohrium (262)	108 **Hs** Hassium (265)	109 **Mt** Meitnerium (266)	110 (269)	111 (272)	112 (277)	113	114				

58 **Ce** Cerium 140.116	59 **Pr** Praseodymium 140.90765	60 **Nd** Neodymium 144.24	61 **Pm** Promethium (145)	62 **Sm** Samarium 150.36	63 **Eu** Europium 151.964	64 **Gd** Gadolinium 157.25	65 **Tb** Terbium 158.92534	66 **Dy** Dysprosium 162.50	67 **Ho** Holmium 164.93032	68 **Er** Erbium 167.26	69 **Tm** Thulium 168.93421	70 **Yb** Ytterbium 173.04	71 **Lu** Lutetium 174.967
90 **Th** Thorium 232.0381	91 **Pa** Protactinium 231.03588	92 **U** Uranium 238.0289	93 **Np** Neptunium (237)	94 **Pu** Plutonium (244)	95 **Am** Americium (243)	96 **Cm** Curium (247)	97 **Bk** Berkelium (247)	98 **Cf** Californium (251)	99 **Es** Einsteinium (252)	100 **Fm** Fermium (257)	101 **Md** Mendelevium (258)	102 **No** Nobelium (259)	103 **Lr** Lawrencium (262)

Figure 1.6. The Periodic Table showing the elements (shaded) that bind to one or both of the metal-binding sites of transferrins.

whereas human lactoferrin (hLf) has a further disulphide bridge in the C-lobe and hSTf has a further four bridges, two in the N-lobe and two in the C-lobe.

Since the earliest report of the structure of transferrins in 1979, from a 6Å electron density map (Gorinsky *et al,* 1979), the tertiary structure of these proteins has been elucidated at much higher resolution. For example, the crystal structure of holo hen ovotransferrin N-lobe has been elucidated recently at 1.65Å (Mizutani *at al,* 2001), human apolactoferrin at 2.0Å (Jameson *et al,* 1998) and rabbit serum transferrin at 2.6Å (Hall *et al,* 2001). However, crystal structures of transferrins complexed to metals other than iron(III) are limited to hLf with copper(III)(Smith *et al,* 1992) and cerium(IV)(Baker *et al,* 2000), and mare lactoferrin with samarium(III)(Sharma *et al,* 1999). These maps show that the transferrins are folded into two lobes of approximately 330 amino acids that are joined together by a short connecting peptide and also make contact via a 'cushion' of hydrophobic residues packed between them. Each lobe can be further split into two dissimilar domains (CI, CII, NI and NII) (Figure 1.7). These domains consist of mixed right-handed β-sheets (~25%) and α-helices (~40%).

Most of the transferrins, except for *Xenopus laevis* transferrin, contain a varying number of N-linked glycosylation sites. For example, rabbit serum transferrin has only one (Evans *et al,* 1988), bovine lactoferrin has four (Baker and Lindley, 1992) and hSTf has two (N413 and N611) (Jamieson *et al,* 1971). No O-linked glycosylations have been identified on transferrins. The fact that not all transferrins are glycosylated, and those which do, have glycosylation sites located at different positions, suggests that there is no specific role for these sugars.

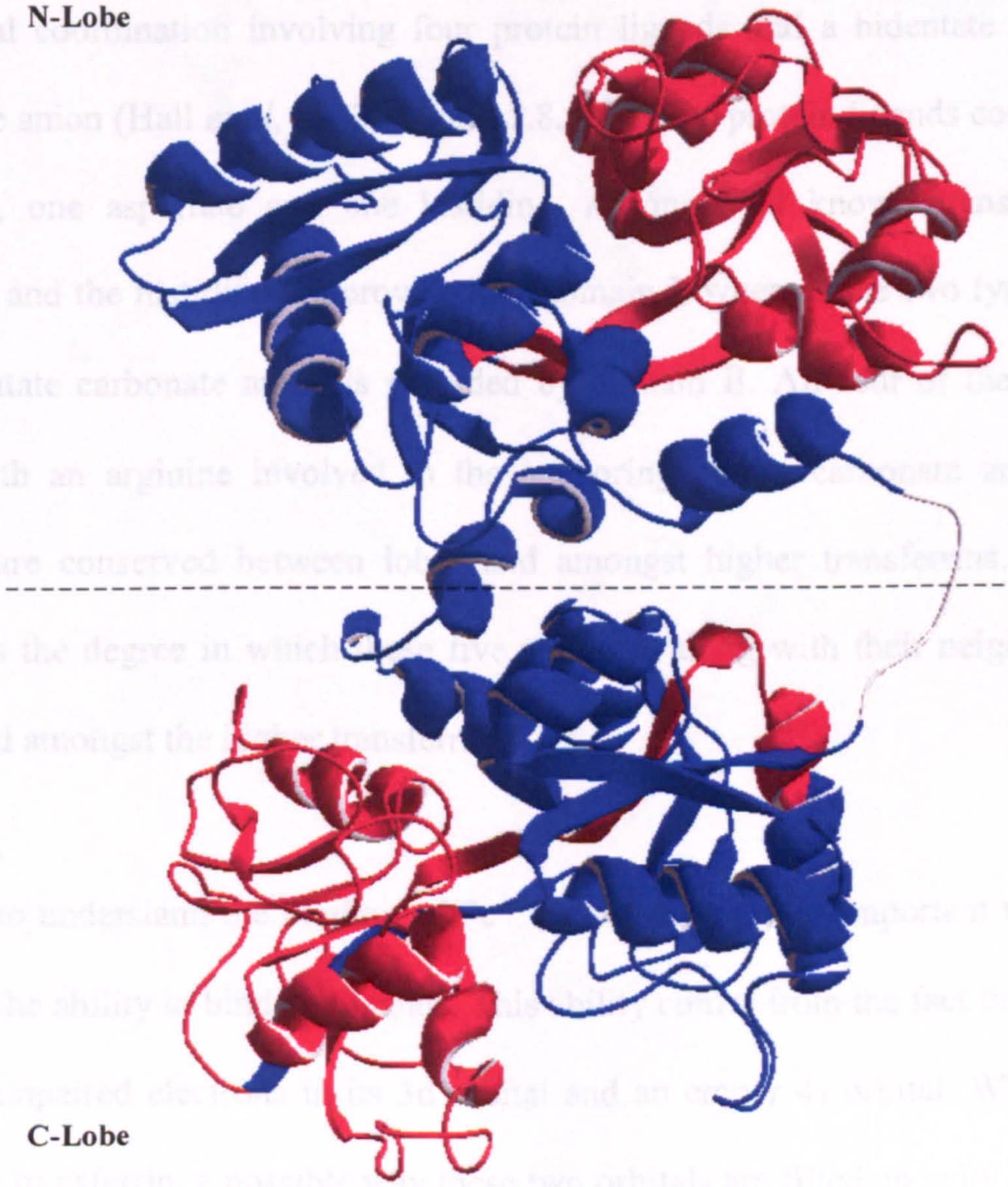


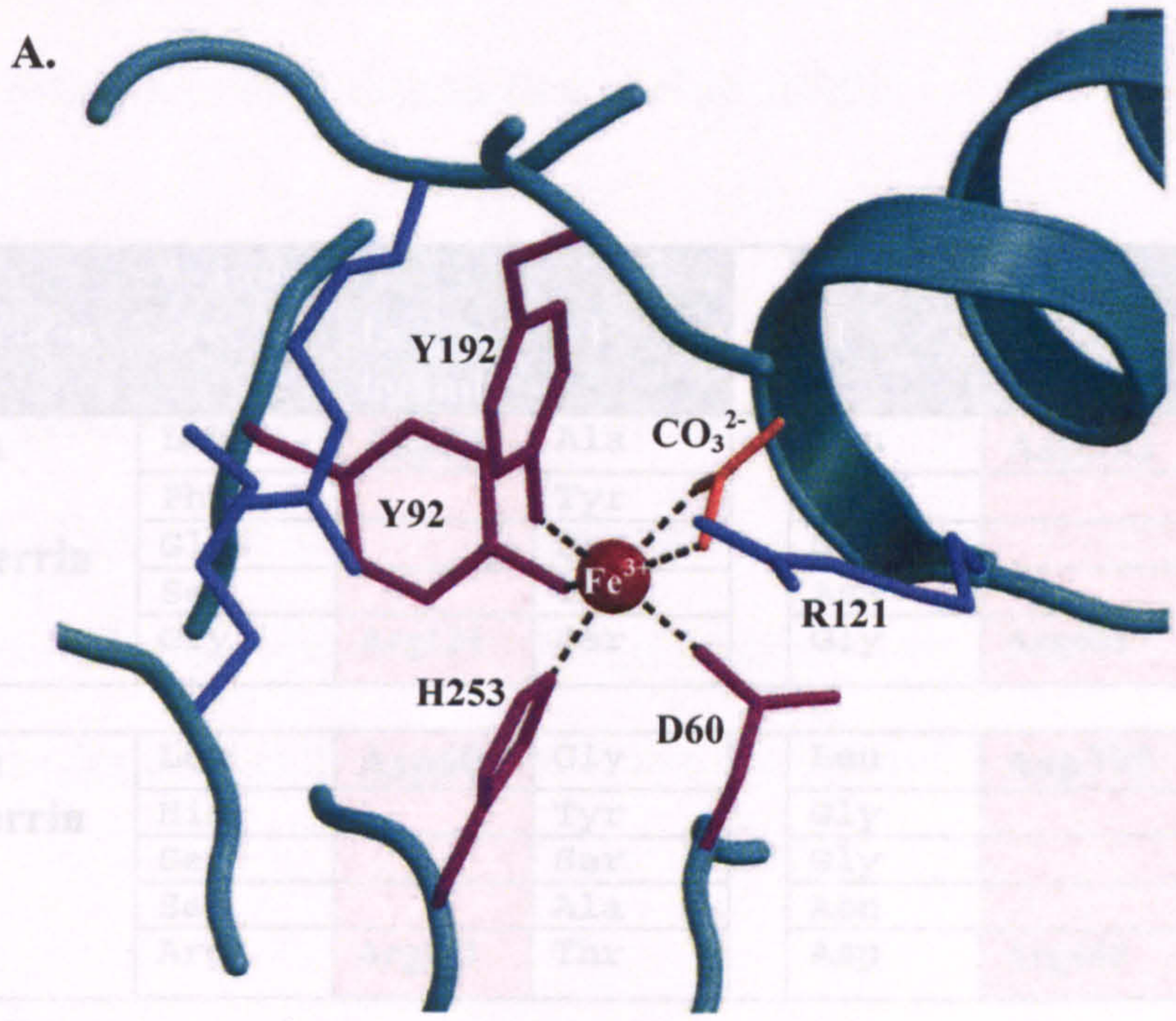
Figure 1.7. Ribbon diagram of rabbit serum transferrin (1JNFA). Picture created using Deep Viewer / Swiss PDB Viewer. Both lobes are labelled and both domains highlighted. Domain I: Blue and Domain II: Red.

1.3.3. Iron-binding Sites

As described previously, transferrins have the capability of reversibly binding two ferric ions, one in each lobe. The two iron-binding sites are 42Å apart and buried deep in the interdomain cleft in a highly polar environment (Anderson *et al*, 1987).

The iron-binding sites in both lobes are very similar and comprise a distorted octahedral coordination involving four protein ligands and a bidentate synergistic carbonate anion (Hall *et al*, 2002)(Figure 1.8.A). These protein ligands comprise two tyrosines, one aspartate and one histidine. Amongst all known transferrins the aspartate and the histidine are provided by domain I, whereas the two tyrosines and the bidentate carbonate anion is provided by domain II. All four of these ligands, along with an arginine involved in the anchoring of the carbonate anion to the protein, are conserved between lobes and amongst higher transferrins. Table 1.4 illustrates the degree in which these five residues, along with their neighbours, are conserved amongst the higher transferrins.

In order to understand the binding of Fe^{3+} by transferrin, it is important to note that Fe^{3+} has the ability to bind six ligands. This ability comes from the fact that the atom has five unpaired electrons in its 3d orbital and an empty 4s orbital. When Fe^{3+} is bound by transferrin, a possible way these two orbitals are filled up is by an electron from one carboxylate oxygen (aspartate), an electron from each of the two phenolate oxygens (two tyrosines), an electron from each of the two oxygens from the anion, and one electron from the imidazole nitrogen (histidine) (Figure 1.8.B). It is also important to note that the three positive charges of the Fe^{3+} are matched by the 3 negative charges of the aspartate and the two tyrosines. Also, the charge on the anion



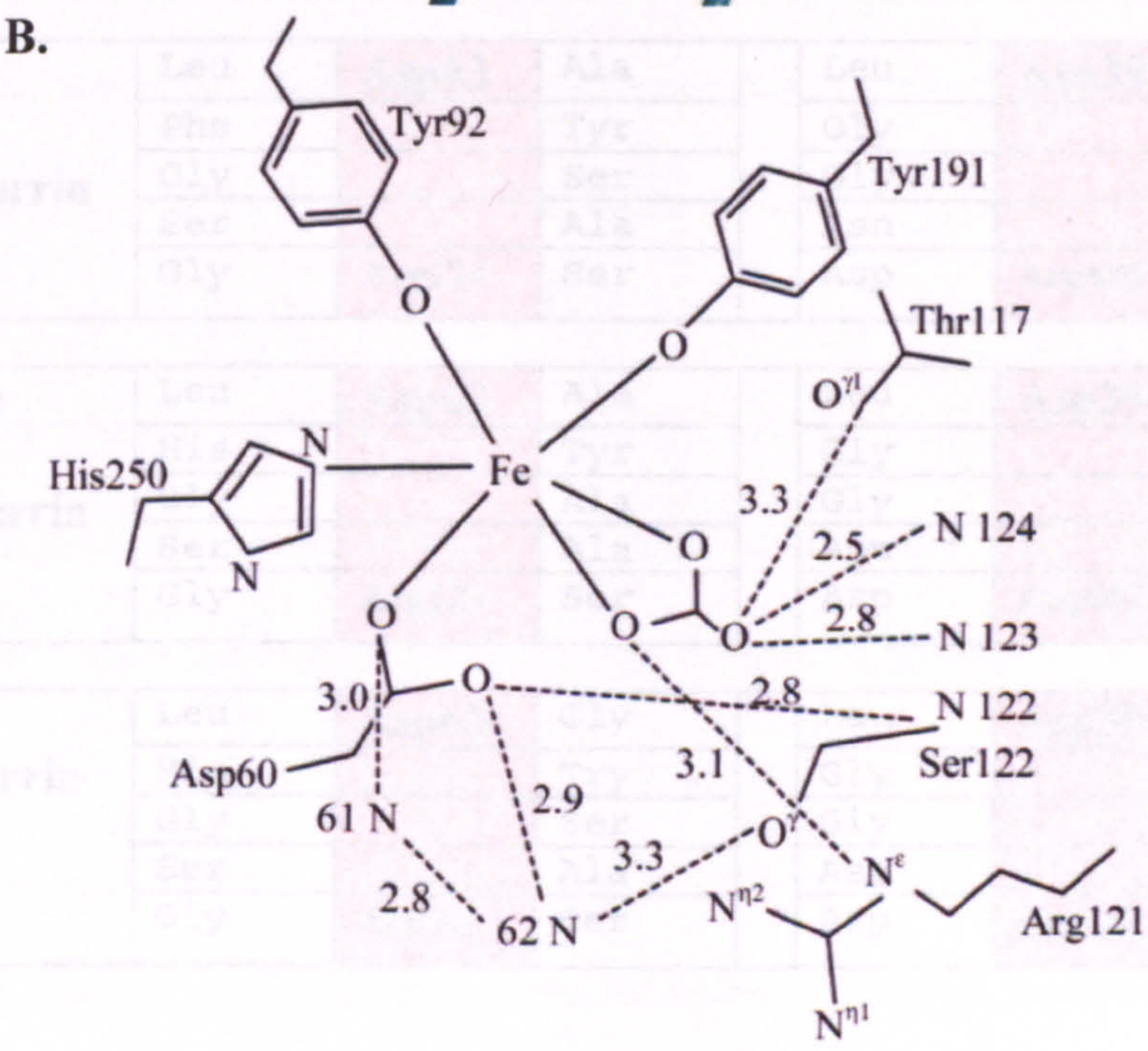


Figure 1.8. A. The N-lobe binding-site of hLf (Baker *et al*, 2002) and B. Schematic diagram of the coordination environment and the hydrogen bonding network around Fe^{3+} in the N-lobe of hen serum transferrin (Guha Thakurta *et al*, 2003). The Fe^{3+} ion is bound to the side chains of Asp60, Tyr92, Tyr192, and His253 and the bidentate CO_3^{2-} ion, which is hydrogen-bonded to Arg121 and to peptide NH groups from the N-terminus of an α-helix.

Transferrin	C-Lobe				N-Lobe		
	-1	Fe^{3+} ligand	+1		-1	Fe^{3+} ligand	+1
Human Serum Transferrin	Leu	**Asp63**	Ala		Leu	**Asp392**	Gly
	Phe	Tyr95	Tyr		Gly	Tyr426	Phe
	Gly	Tyr188	Ser		Gly	Tyr517	Thr
	Ser	His249	Thr		Asn	His585	Ala
	Gly	**Arg124**	Ser		Gly	**Arg456**	Thr
Human Lactoferrin	Leu	**Asp60**	Gly		Leu	**Asp395**	Gly
	His	Tyr92	Tyr		Gly	Tyr435	Leu
	Ser	Tyr192	Ser		Gly	Tyr528	Thr
	Ser	Tyr253	Ala		Asn	His597	Ala
	Arg	**Arg121**	Thr		Asp	**Arg465**	Thr
Rabbit Serum Transferrin	Leu	**Asp63**	Ala		Leu	**Asp392**	Gly
	Phe	Tyr95	Tyr		Gly	Tyr425	Leu
	Gly	Tyr188	Ser		Gly	Tyr514	Thr
	Ser	Tyr249	Ala		Asn	His582	Ala
	Gly	**Arg124**	Ser		Asp	**Arg455**	Thr
Porcine Serum Transferrin	Leu	**Asp62**	Ala		Leu	**Asp396**	Gly
	His	Tyr94	Tyr		Gly	Tyr431	Leu
	Gly	Tyr192	Ala		Gly	Tyr526	Thr
	Ser	His253	Ala		Asn	His594	Ala
	Gly	**Arg123**	Ser		Asp	**Arg462**	Thr
Camel Lactoferrin	Leu	**Asp60**	Gly		Leu	**Asp395**	Gly
	His	Tyr92	Try		Gly	Tyr433	Leu
	Gly	Tyr192	Ser		Gly	Tyr526	Thr
	Ser	His253	Ala		Asn	His595	Ala
	Gly	**Arg121**	Ser		Asp	**Arg463**	Thr

Table 1.4. Comparison of residues involved in iron binding, their position and the amino acid either side of it (-1 and +1) for both lobes amongst different transferrins. Residues highlighted in red are those directly involved in the binding of the ferric ion. Out of these residues, yellow text represents the residues involved in carbonate binding.

is compensated for by the positively charged arginine side chain and a positive charge at the N-terminus of an α-helix (Baker *et al*, 2003b).

1.3.4. Iron Binding and Release

A feature common to all transferrins is the closing of the interdomain clefts upon incorporation of iron into the iron-binding sites, in doing so, bringing the iron-binding ligands on both domains together, 'grasping' the iron. This feature has been shown to be the opposite in the apo-form, in that the domains are open (Baker *et al*, 1997).

The actual mechanism of iron binding is still not completely clear. It is thought that iron-binding starts with the binding of the anion to domain II, followed closely by the binding of the metal by the two tyrosines on the same domain. Once bound, a conformational change occurs, causing domain I to close over domain II. This closure allows the histidine and aspartate ligands to come into contact with the iron, allowing them to bind the iron's remaining two coordinates.

The extent in which the two domains open and close, along with how highly conserved the rotation angle of each of the domains are across the transferrin family (Guha Thakurta *et al*, 2004). They showed that if domain I, from the closed form of hen STf N-lobe, was superimposed with the open form, a rotation of 52.7° difference can be seen. Similarly in the C-lobe, a rotation of 35° was seen. The rotation in the N-lobe of human lactoferrin and human serum transferrin is also similar, 54° and 63°, respectively.

The actual affinity for the Fe^{3+} ion varies between both lobes and also members of the transferrin family. One explanation for this is the iron-ligand bond distances, as these distances are known to vary. The longer bond length is thought to be the reason why the N-lobe of serum transferrin releases iron more readily than the C lobe. Iron is released from the N-lobe of serum transferrin at around pH 5.7, whereas the C lobe retains iron until around pH 4.8 (Rinaldo and Field, 2003). This difference can also be seen between transferrin family members as, for example, lactoferrin retains iron at a pH as low as 3.5. The higher iron-binding affinity seen in lactoferrins could also be due to the greater number of interdomain hydrogen bonds (Hall *et al,* 2002).

The differences in iron affinity between the serum transferrins and the lactoferrins has also been proposed to be due to regions found in the N-terminal lobe of only the serum transferrins, away from the iron-binding site, at the interdomain interface. Around this region, two lysines (K206 + K296) have been proposed to play a role in iron-release. These residues are on opposite faces of the iron-binding cleft and are thought to be bridged by a single proton when the protein is in its closed form. The theory is, at low pH (section 1.2.3) these two lysines are protonated, leading to the subsequent release of the proton bridge and causing the two domains to repel each other (Rinaldo and Field, 2003). As the domains flip open, the iron is released. This theory is known as the dilysine trigger. The importance of these lysines were highlighted by Nurizzo *et al* in 2001 when they showed that mutating these two lysines had a significant effect on iron-binding and release.

However, a primary role of this dilysine trigger has been questioned. It is thought that the protonation of the carbonate anion and the histidine iron-binding ligand

constitutes the primary stages of iron release, with the dilysine trigger being important once the iron binding ligands are weakened (Hall *et al,* 2002). However, this dilysine trigger does explain, at least in part, the *in vitro* release of iron more readily from the N-lobe compared to the C-lobe of serum transferrin.

1.3.5. Tf/TfR Interactions

As described earlier (section 1.2.3), serum transferrin transports iron from the gut and delivers it to cells using an endocytotic pathway involving the transferrin receptor (TfR). At the slightly alkaline pH of 7.4, two diferric serum transferrin molecules can bind to the dimeric transferrin receptor. A model of the Tf/TfR complex is shown in Figure 1.9. This figure shows the orientation in which the two diferric Tf molecules bind to the TfR. The model also shows the N-lobe to bind beneath the receptor, making contact with the protease-like and helical domains, therefore being sandwiched between the membrane and the TfR ectodomain, and the C-lobe pointing up towards the helical domain (Cheng *et al*, 2004). On binding, there also seems to be a shift of the N-lobe by about 9Å with respect to the C-lobe, 'stretching' the protein onto the receptor, tightening the interactions. Regions on the TfR shown to interact in this model are consistent with regions predicted through mutagenesis studies (Giannetti *et al*, 2003).

In the C-lobe a large interaction is thought to take place between domain I and the receptor, leaving domain II to move freely. This would suggest that iron could be released from the C-lobe whilst the transferrin is still attached to the receptor. The interaction of the N-lobe to the receptor seems to be somewhat different. This lobe has been found to make two very localised interactions with the receptor. Domain I is

Figure 1.9. An atomic model of the TfR diferric Tf complex created by Cheng *et al*, 2004. The model was created by fitting the crystal structure of the TfR with the crystal structure of the human serum Tf N-lobe (MacGillivray *et al*, 1998) and the crystal structure of rabbit serum Tf (Hall *et al*, 2002).

thought to interact with the helical domain of the receptor and domain II is thought to interact with the protease-like domain. Even though both domains in the N-lobe are involved in receptor binding, studies have shown that the N-lobe interacts with the TfR more weakly than the C-Lobe. Cheng *et al*, 2004 also showed that the N-lobe has a hydrophobic patch that is thought to make it non-specifically 'sticky', giving it the ability to bind to different regions of the receptor.

Receptor bound Tf preferentially releases iron from the C-lobe, probably due to the freedom of movement of domain II. Cheng *et al*, 2004 have shown the C-lobe to rotate approximately parallel to the membrane on release of the iron. However, because the N-lobe interacts with the receptor using both of its domains, it cannot release the iron so easily. This could explain the predominance of N-monoferric Tf in the serum. However, lobe communication through the bridge connecting the two lobes and the interaction of the C-terminal helix with domain I of the N-lobe, is thought to be a major key in the release of iron from both lobes and the Tf from the receptor. Gumerov *et al*, 2003, confirm this theory of inter-lobe communication as they show that loading of the C-lobe with iron reduces the conformational flexibility of the iron-free N-lobe. They then postulate that this in turn could cause a transducer effect whereby the receptor could undergo a conformational change and thus forcing a transition of Tf from its closed conformation to its open one. Evans and Williams also reported this inter-lobe communication in 1980, when they found that C-lobe monoferric hSTf had a more stable N-lobe than the N-lobe of apo-hSTf.

1.3.6. Transferrin-Iron Uptake by Bacteria

The ability to sequester iron by bacteria is recognised as a virulence factor since the environment in a mammalian host is believed to be iron limited. Mechanisms for this iron acquisition fall into two general categories: synthesis of siderophores and cognate receptors, or receptor mediated acquisition of iron from host proteins such as transferrin (Ratledge and Dover, 2000).

A number of bacteria are capable of utilising hSTf- and hLf- bound iron. These include pathogenic *Neisseria*, *Moraxella catarrhalis*, *Actinobacillus pleuropneumoniae*, *Haemophilus influenzae* and *Pasteurella multocida* (Cornelissen, 2003). Iron uptake from hSTf is mediated by two proteins, transferrin-binding proteins A and B (TbpA and TbpB), while lactoferrin-binding proteins A and B (LbpA and LbpB) are associated with iron-uptake from hLf.

Figure 1.10 shows a schematic representation of the TbpAB/LbpAB system involved in iron uptake from transferrins by *Neisseria meningitidis.* The mechanism of iron acquisition is thought to start with the initial binding of TbpB/LbpB with diferric transferrin, followed with the binding of this protein complex with TbpA/LbpA. Once all three proteins are complexed, the diferric transferrin is in a position whereby it sits over the TbpA/LpbA. Once here, iron is released from the Tf molecule and passes through TbpA/LbpA. Ferric ion transport through these membrane proteins is thought to be dependent on the TonB-protein. TonB, along with it cytoplasmic membrane accessory proteins ExbB and ExbD, operates in this system as an energy transducer that couples proton-motive force between the two membranes (Gorringe and Oakhill, 2002). Following passage through the outer

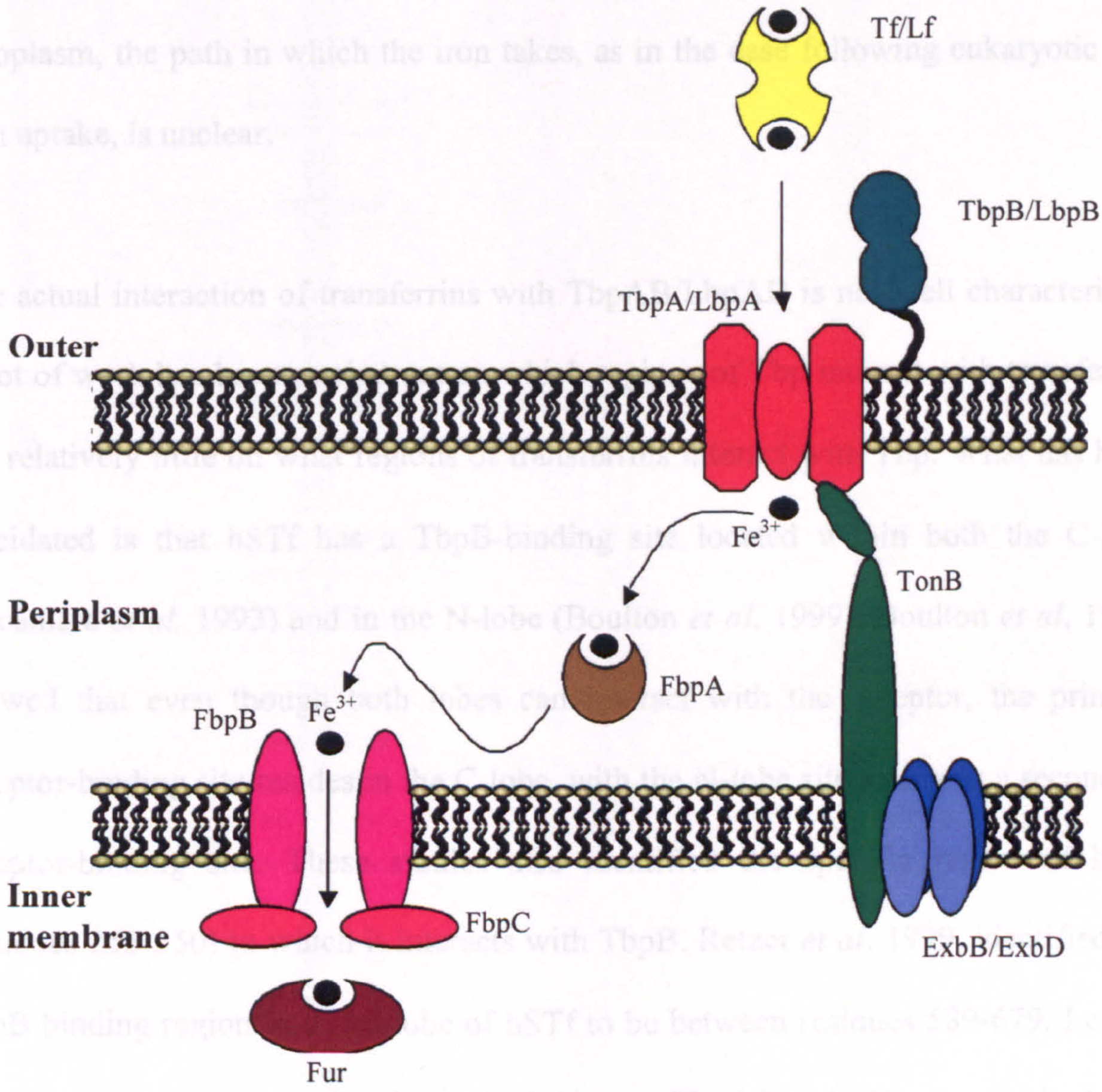


Figure 1.10. A schematic representation of the system involved in iron uptake from transferrins by *Neisseria meningitidis*. Internalisation of iron into the bacterial cell is thought to be mediated by TbpAB/LbpAB using energy derived from TonB. (Adapted from Gorringe and Oakhill, 2002).

membrane, ferric ion is then thought to be taken up by ferric binding protein A (FbpA). This protein then shuttles the iron from the outer membrane to the inner membrane where it interacts with FbpB, a permease. As FbpC possesses an ATP-binding domain, it has been postulated that this protein supplies FbpB with the energy required for active transport of iron into the cytoplasm. Once in the cytoplasm, the path in which the iron takes, as in the case following eukaryotic cell iron uptake, is unclear.

The actual interaction of transferrins with TbpAB/LbpAB is not well characterised. A lot of work has been carried out on which regions of Tbp interact with transferrin, but relatively little on what regions of transferrins interact with Tbp. What has been elucidated is that hSTf has a TbpB-binding site located within both the C-lobe (Alcantara *et al,* 1993) and in the N-lobe (Boulton *et al*, 1999). Boulton *et al*, 1999, showed that even though both lobes can interact with the receptor, the primary receptor-binding site resides in the C-lobe, with the N-lobe site acting as a secondary receptor-binding site. These studies also identified the specific region of hSTf (residues 255-350) in which it interacts with TbpB. Retzer *et al*, 1999, identified the TbpB binding region in the C-lobe of hSTf to be between residues 589-679. Less is known about the regions involved in binding to TbpA/LbpA. However, LbpA has been shown to bind to both domains of the C-lobe of hLf (Wong and Schryvers, 2003).

1.3.7. Genomic Organisation of hSTf and hLf

Both hSTf and hLf are located within close proximity from each other in the genome, 3q22.1 and 3q21-q23, respectively (www.ncbi.nlm.nih.gov/genome). Figure 1.11

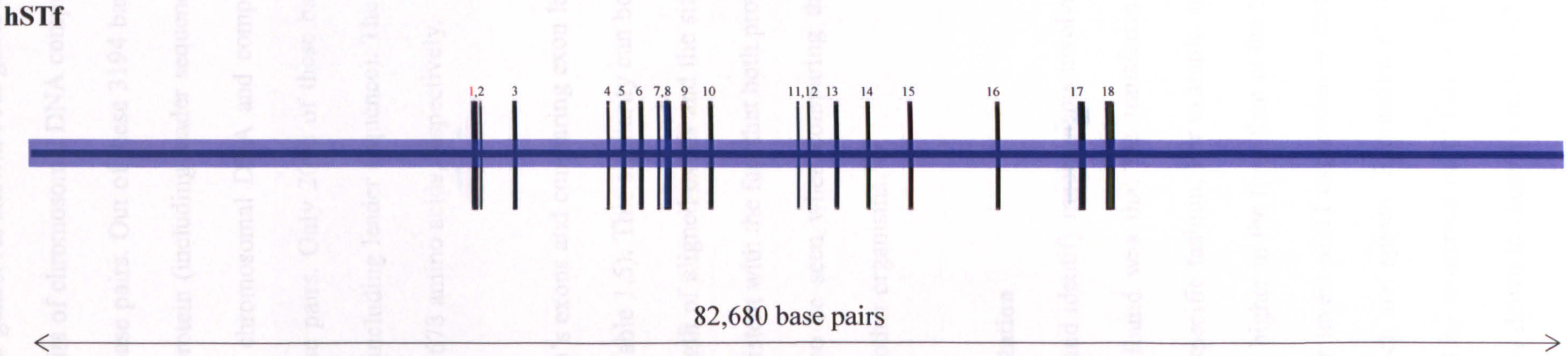


Figure 1.11. A schematic representation of the hSTf and the hLf gene. All exons are numbered above their relative position and all introns are highlighted in blue. Exon numbers coloured red indicate exons which are not involved in coding for the protein (data from www.ncbi.nlm.nih.gov/genome).

shows the genomic map of both hSTf and hLf genes. As shown, both genes are quite different. The hLf gene spans 73692 base pairs of chromosomal DNA comprising of 21 exons, with a total exon length of 3194 base pairs. Out of these 3194 base pairs, 2136 base pairs (17 exons) encode for the protein (including leader sequence). The hSTf gene spans over 82676 base pairs of chromosomal DNA and comprises 18 exons, with a total exon length of 2664 base pairs. Only 2096 of these base pairs (spanning 17 exons) encode for the protein (including leader sequence). The sizes of the mature hLf and hSTf protein are 691 and 678 amino acids, respectively.

Following the alignment of these two protein's exons and comparing exon lengths, a high degree of similarity can be observed (Table 1.5). This similarity can be seen in the number of protein coding exons, the length of aligned exons and the size of the leader sequences. These similarities are consistent with the fact that both proteins are related. However, these similarities can also be seen when comparing these two genes with the genes encoding transferrin in other organisms.

1.3.8. Transferrin Gene Expression Regulation

Several studies have been carried out to try and identify mechanisms involved in the regulation of hSTf expression. What was found was that the transferrin gene is induced by many different factors in tissue specific fashion. For example, in the rat, the levels of transferrin mRNAs are 6.5-fold higher in the liver than in the testis and the brain (Idzerda *et al*, 1986). Much lower levels of Tf expression is observed in many other tissues such as the heart, stomach and spleen. Expression regulation is not just tissue specific, but is also regulated by numerous other factors. In rats and chickens, nutritional iron deficiency has been shown to cause a specific induction of

hLf		hSTf	
Exon	Coding Exon Length *(No. amino acids)*	Exon	Coding Exon Length *(No. amino acids)*
1		-	
2		-	
3		-	
4		1	
5	14	2	14
6	5, 50	3	5, 53
7	36	4	36
8	61	5	59
9	50	6	45
10	18	7	18
11	60	8	60
12	58	9	59
13	52	10	52
14	31	11	31
15	18	12	11
16	52	13	52
17	48	14	46
18	22	15	21
19	62	16	62
20	63	17	63
21	11	18	35

Table 1.5. Alignment of exons and comparing coding exon length (amino acids) between hLf and hSTf. Numbers in blue represent the protein's leader sequence. Exons highlighted in yellow represent exons of the same length (data from www.ncbi.nlm.nih.gov/genome).

transferrin gene transcription in the liver, resulting in a several fold increase in transferrin synthesis (Idzerda *et al,* 1989). Glucocorticoids and oestrogen have also been shown to stimulate hepatic transferrin synthesis. In chicken, these hormones are known to act at the transcriptional level to increase hSTf synthesis by roughly two-fold.

Idzerda *et al*, 1989, showed, through deletion analysis of a 5 kb pairs region upstream of the 5' end of the transferrin gene, four major functionally different regions (Figure 1.12.A). These regions include a cell-type specific promoter localised between positions –145 and –45 base pairs, a distal promoter from –620 to –125 base pairs containing positive and a negative *cis*-acting elements, a negative acting region between –1.0 and –0.6 Kb pairs which functions as a silencer (Ghareeb *et al*, 1998) and a strong enhancer located between –4.0 and –3.3 kb pairs which is known to interact with hypoxia-inducible factor-1 (Vyhlidal and Safe, 2002). More specifically, the region from +1 to –600 base pairs can be expanded further to include a number of different protein-binding regions (Figure 1.12.B). Schaeffer *et al*, 1993, showed that two of these regions include two proximal promoters (PRI and PRII). These two are known to make up part of the tissue-specific promoter. For example, in the liver, hepatocyte nuclear factor 4 acts at the PRI site, while CCAAT/enhancer-binding proteins act at the PRII site. In the testis, Sertoli proteins SP-A and SP-D binds to the PRBI site, while SP-α and SP-β bind to the PRII site. These two sites share a common 10-nucleotide-long inverted repeated sequence with two of the other regulatory regions known to make up the distal promoter region (DRI and DRII) (Brunel *et al*, 1988). However, identities of proteins that bind DRI

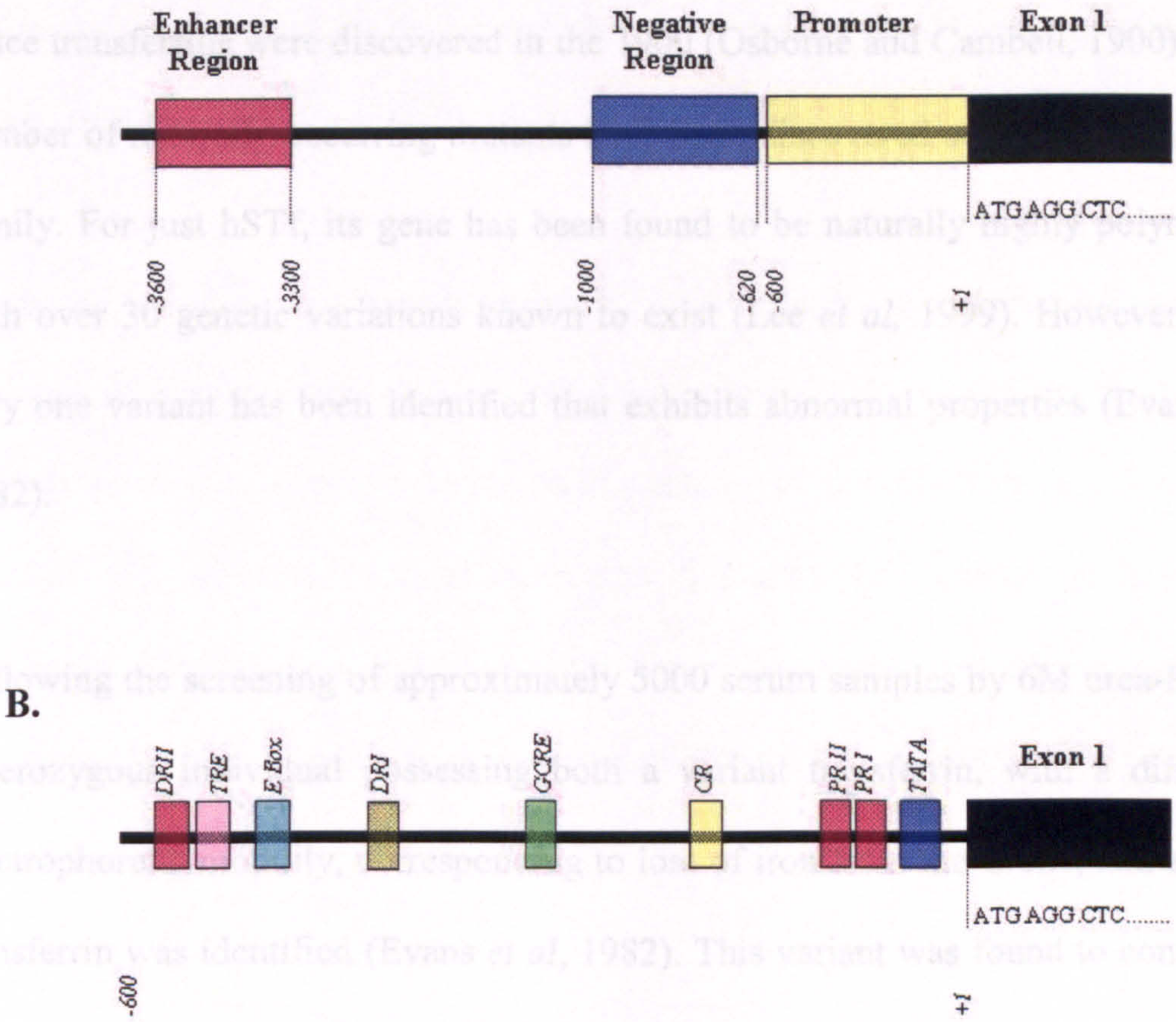


Figure 1.12. A schematic representation of the regulatory components upstream of the hSTf gene that are thought to be in involved in regulation of expression. A: Regions that have been linked to transferrin gene expression upstream of 5' end. B: Protein-binding sites in the promoter region. *(Modified from Testa, 2002).*

and DRII are unknown. A fifth region, known as a central regulatory region (CR), has also been identified (Schaeffer *et al*, 1993), but as yet is poorly characterised.

1.3.9. Transferrin Mutants

Since transferrins were discovered in the 1900 (Osborne and Cambell, 1900), a large number of naturally occurring mutants have been discovered across the transferrin's family. For just hSTf, its gene has been found to be naturally highly polymorphic, with over 30 genetic variations known to exist (Lee *et al,* 1999). However, so far, only one variant has been identified that exhibits abnormal properties (Evans *et al,* 1982).

Following the screening of approximately 5000 serum samples by 6M urea-PAGE, a heterozygous individual possessing both a variant transferrin, with a differential electrophoretic mobility, corresponding to loss of iron from the C-site, and a normal transferrin was identified (Evans *et al*, 1982). This variant was found to contain a G to A substitution in the first nucleotide of codon 394, leading to the mutation G394R (Evans *et al*, 1994). This variant was shown to have a reduced affinity for iron in the C-lobe and also a reduced affinity for the transferrin receptor (Young *et al,* 1984).

Evans *et al*, 1994, suggested that G394 is not directly involved in iron-binding, nor does it seem likely to be involved in receptor binding given that it is buried deep within the cleft. In 1994, Evans *et al* predicted that the effect of G394R on iron binding was due to its interaction with the iron-binding residue D392, preventing it from interacting with the iron. They showed that the inability of the aspartate ligand to bind to iron is consistent with the observed blue shift in the visible spectrum of the

C-site as judged by qualitative comparison with the mutants of D63 or the N-lobe protein (Woodworth *et al,* 1991) and with an 18-kDa yellow fragment of duck ovotransferrin (Evans and Madden, 1984), which is known to lack the aspartate ligand. However, even though their discussion seemed to be conclusive, the reason for the reduced affinity for the receptor has yet to be elucidated, and the reduced iron binding affinity is still unclear.

Since the advent of recombinant protein technology, a number of genetically engineered transferrin mutants have also been created. The reason for creating such variants is to try to gain a better understanding of the structure and function of transferrins by mutating residues and regions thought to be important in the structure and function of the molecule. Table 1.6 lists a few examples of these genetically engineered mutants, along with a few examples of the naturally occurring ones.

Mutation	Description	Reference
Native		
G277S	A risk factor for iron deficiency anaemia	Lee *et al*, 2001
D277G	Four non-symptomatic hSTf variants	Lee *et al*, 1999
H300R		
G652E		
P570S		
K627E	One non-symptomatic hSTf variant	Pang *et al*, 1998
G394R	Variant of hSTf with abnormal functional properties	Evans *et al*, 1982
Recombinant		
D63S	Non-glycosylated transferrin mutants expressed in Baby Hamster Kidney (BHK) cells	Mason *et al*, 1998
D392S		
D63E	N-terminal half transferrin mutant expressed in BHK cells	Baker *et al*, 2003
R124A		
K206A	N-terminal half transferrin mutants expressed in *Pichia pastoris*	Steinlein *et al*, 1998
K296A		
K206Q	N-terminal half transferrin mutant expressed in BHK cells	Yang *et al*, 2000
H207E		
H249A	N-terminal half transferrin mutant expressed in BHK cells	He *et al*, 2000
H249E		
H249Q		
M109L	N-terminal half transferrin mutant expressed in BHK cells	He *et al*, 1999
W8Y	N-terminal half transferrin mutant expressed in BHK cells	He *et al*, 2001
W128Y		
W264Y		
Y85F	N-terminal half transferrin mutant expressed in BHK cells	He *et al*, 1998
E83A		

Table 1.6. Examples of naturally occurring hSTf mutants and some genetically engineered mutants.

1.4. Aims and Objectives

As described previously in this chapter, hSTf provides not only a means of transporting iron from the sites of absorption and storage to the sites of utilisation, but is also a protection against the damaging effects of iron-catalysed free radical cascades. Its role in iron homeostasis is therefore essential, but although the protein was identified more than half a century ago, its full-length crystal structure and its complete mechanism of iron-loading and releasing, including receptor interaction, is still yet to be elucidated.

The aim of this thesis is therefore to look deeper into the structure and function of hSTf in the hope to gain a better understanding of not only this specific protein, but also of the other members of the transferrin family and the whole regulatory network of iron metabolism. The scheme of this study is summarised in Figure 1.13.

The initial aim of this project is to create an expression system for the synthesis of full-length hSTf and hLf. Recombinant protein technologies are now one of the main tools in research used to gain a better understanding of proteins, enabling the identification of functionally and structurally important residues, or regions, in the protein. Expression of such proteins enables one to modify the protein in many ways, either in the form of specific amino acid mutations, protein region deletions, creation of protein-protein fusions, or creating non-glycosylated proteins. In these studies, the aim is to concentrate on specific amino acid mutations in and around the iron binding-site and theoretical receptor binding sites. Such mutations will include the amino acid substitution at position 394 (Gly → Arg) of hSTf, in a naturally occurring variant first identified in 1982 (Evans *et al,* 1982).

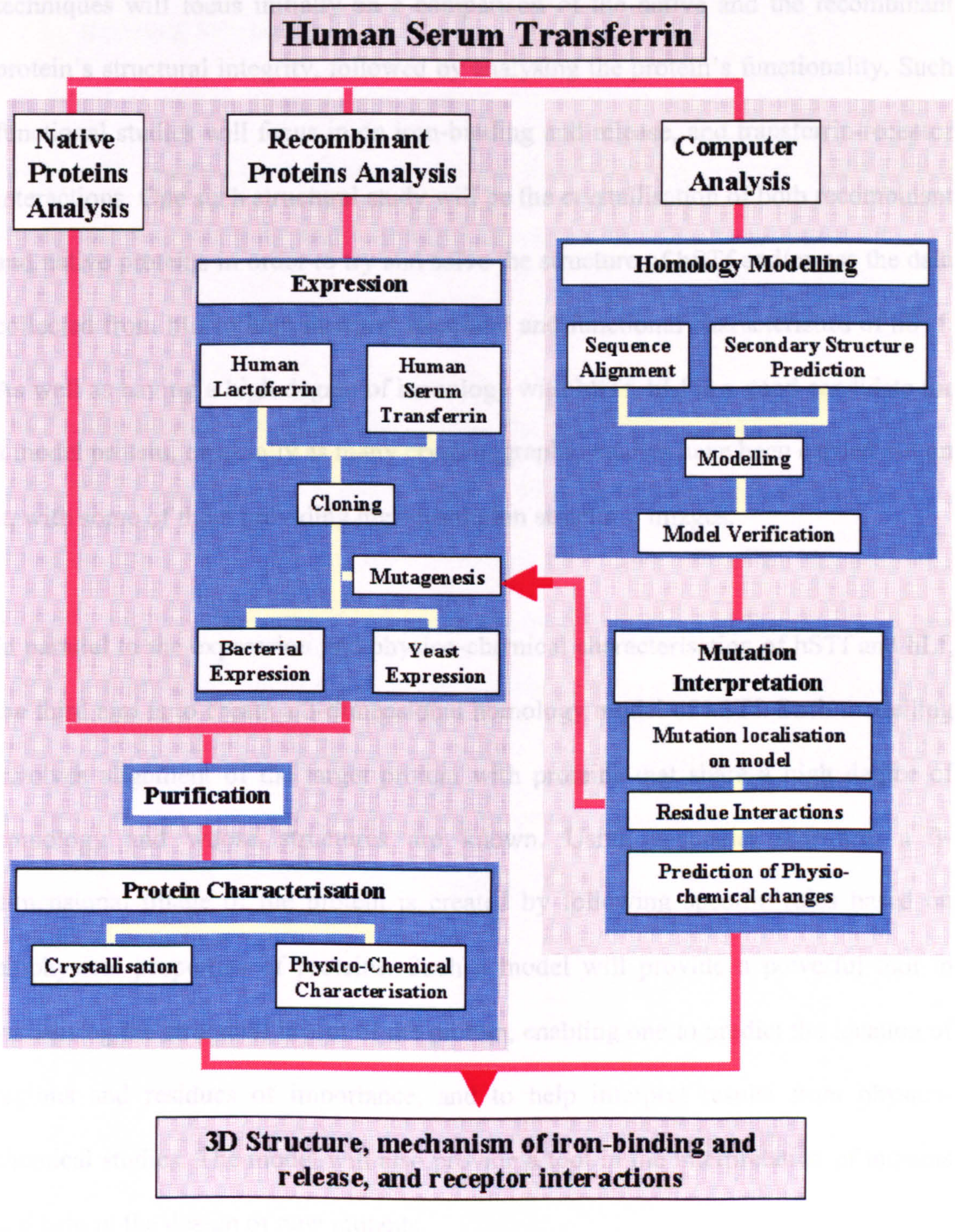


Figure 1.13. Plan of study. The aims are split into protein analysis and computer analysis, with the computer analysis results influencing the mutagenesis studies.

Once purified, the second aim is to analyse these recombinant proteins, along with their native counterparts using a variety of physico-chemical techniques. These techniques will focus initially on a comparison of the native and the recombinant protein's structural integrity, followed by analysing the protein's functionality. Such functional studies will focus in on iron-binding and release, and transferrin-receptor interactions. One such structural study will be the crystallisation of both recombinant and native proteins in order to try and solve the structure of hSTf and/or use the data collected from hLf to help interpret structural and functional characteristics of hSTf. As well as having a high degree of homology with hSTf, hLf is a good candidate for a model protein, especially as many crystallographic studies have been carried out on it, with some of them providing high-resolution structural images.

In parallel to the expression and physico-chemical characterisation of hSTf and hLf, the third aim is to create a 3-dimensional homology model of hSTf. Such modelling involves alignment of the target protein with proteins that share a high degree of homology and whose structures are known. Using specialist software, a 3-dimensional image of the protein is created by following specific rules based on biophysical properties of proteins. Such a model will provide a powerful tool in visualising the structural layout of the protein, enabling one to predict the location of regions and residues of importance, and to help interpret results from physico-chemical studies. The model will also provide a tool in the interpretation of mutants and help in the design of new mutants.

2. Materials and Methods

2.1. Materials

2.1.1. Bacterial Strains

The bacterial strains used in these studies are listed in Appendix II.1. All *E.coli* strains were maintained in LB medium containing 15% (v/v) glycerol at -70°C. For day-to-day use, cells were maintained on LB agar medium plates and kept at 4°C.

2.1.2. Yeast Strains

The *Saccharomyces cerevisiae* cells used for the expression of hSTf were derivatives of the DXY1 strain (Kerry-Williams *et al,* 1998). This strain contains mutations in two of its genes, *LEU2* and *YAP3*. *LEU2* encodes an enzyme involved in the leucine synthesis pathway. Therefore, cells with a mutation in the *LEU2* gene *(leu2)* are unable to grow in the absence of leucine, allowing for the selection of cells that have been transformed with yeast expression vectors containing the intact *LEU2* gene. The disruption in the *YAP3* gene (*yap3::LYS2*) prevents the expression of a yeast aspartyl protease previously shown to proteolytically cut recombinant human albumin if expressed by the cell.

2.1.3. Microbiological Media

The different media used in these studies is shown in Table 2.1. Bacto-agar, yeast extract, yeast nitrogen base (w/o amino acids and ammonium sulphate), bacto-tryptone and bactopeptone were purchased from Difco Laboratories. All antibiotics were purchased from Sigma Aldrich, together with X-gal and IPTG. The antibiotics used in these studies are shown in Table 2.2.

Table 2.1. Name and composition of different media used in these studies.

Media Name	Content
Luria Bertani (LB)	10g/L bacto-tryptone, 10g/L sodium chloride, 5g/L yeast extract, adjusted to pH 7.5 with NaOH
SOC	To 950mL MilliQ water, 20 g bacto-tryptone, 5g bacto-yeast extract and 0.5g NaCl were added. After dissolving the ingredients, 10mL of a 250mM solution of KCl was added and the pH of the medium was adjusted to pH 7 with NaOH. The volume of the medium was made up to 1L with milliQ water. Just before use, 5mL of sterile 2M $MgCl_2$ and 20mL of sterile 1M glucose was added.
BMM-D	1.7g/L Yeast Nitrogen Base (w/o amino acids and ammonium sulphate), 5g/L ammonium sulphate, 36mL/L of 1M Citric Acid, 126mL/L of 1M di-sodium hydrogen orthophosphate and 2% (v/v) D-glucose.
YEPD	10g/L yeast extract, 20g/L bacto-peptone and 2% (v/v) glucose

Agar plates were made by the addition of 15g/L of Bacteriological agar to the specific media followed by autoclaving.

Table 2.2. Name and working concentration of antibiotics used in these studies.

Antibiotic	Stock Solution	Working Concentration
Ampicillin	50-100mg/ml	50-100µg/ml
Apramycin	35mg/ml	35µg/ml
Chloramphenicol	20mg/ml	20µg/ml
Tetracycline	10mg/ml	10µg/ml

2.1.4. Chemical Solutions

Unless otherwise stated in the text, all chemicals were supplied by BDH Chemicals Ltd or Sigma Aldrich and were of Analar grade or equivalent.

2.1.5. Buffers

DNA Agarose Gel TBE Buffer (x10) -*Autoclaved, then adjusted to pH ~8.2 – 8.4 using 11.6M HCl*

Component	Final Concentration	Quantity
Tris Base	1M	121.1g/L
Boric Acid	1M	61.83g/L
EDTA	20mM	7.44g/L

Cell Suspension Buffer *–adjusted to pH 8 using 11.6M HCl*

Component	Final Concentration	Quantity
Tris Base	50mM	6.055g/L
EDTA	1mM	372.24mg/L

Citric Acid Buffer *–adjusted to pH 4.7 using 10M NaOH*

Component	Final Concentration	Quantity
Citric Acid	100mM	21.014g/L

Coupling Buffer *–adjusted to pH 6.5 using 10M NaOH*

Component	Final Concentration	Quantity
Citric Acid	200mM	42.028g/L

DEAE Elution Buffer *-pH not adjusted (~pH 9.4)*

Component	Final Concentration	Quantity
Potassium Tetraborate Tetrahydrate	82.6mM	25.2g/L

DEAE Equilibration Buffer 1 - *pH not adjusted (~pH 9.3)*

Component	Final Concentration	Quantity
Potassium Tetraborate Tetrahydrate	11mM	3.3605g/L

DEAE Equilibration Buffer 2 - *pH not adjusted (~pH 9.2)*

Component	Final Concentration	Quantity
Potassium Tetraborate Tetrahydrate	15.7mM	4.8g/L

FeNTA (0.1M) - *NTA dissolved in 2M NaOH first then made up to 10mL with MilliQ water following the addition of Iron(III) Chloride-6-Hydrate.*

Component	Final Concentration	Quantity
Nitrilotriacetic Acid N_1N-bis [Carboxymethyl]glycine	0.1M	191mg
2M Sodium Hydroxide	0.2M	1mL
Iron(III) Chloride-6-Hydrate	0.1M	2mL

Immunoaffinity Buffer *–adjusted to pH 8 using 11.6M HCl*

Component	Final Concentration	Quantity
Tris Base	20mM	2.422g/L
NaCl	100mM	5.844g/L
EDTA	1mM	372.24mg/L

ITC Sample Buffer *–adjusted to pH 8 using 11.6M HCl*

Component	Final Concentration	Quantity
Tris Base	20mM	2.42g/L
Sodium Hydrogen Carbonate	20mM	1.68g/L
Sodium Chloride	50mM	2.92g/L

Native-PAGE Running Buffer (x10) - *Made up to 1litre with MilliQ water. pH not adjusted (~8.8)*

Component	Final Concentration	Quantity
Tris Base	250mM	30.3g/L
Glycine	1.92M	144g/L

Nitrocellulose Blocking Buffer

Component	Final Concentration	Quantity
PBS-T	1 x	250mL
Bovine Serum Albumin	1% (w/v)	2.5g

PBS-T

Component	Final Concentration	Quantity
PBS	1 x	250mL
Polyoxyethylene-sorbitan mono-laurate ($Tween_{20}$)	0.05% (v/v)	125µL

SDS-PAGE Running Buffer (x10) -*Made up to 1litre with MilliQ water. pH not adjusted (~8.3).*

Component	Final Concentration	Quantity
Tris Base	250mM	30.3g/L
Glycine	1.92M	144g/L
Sodium dodecyl Sulphate	35mM	10g/L

SP Elution Buffer -adjusted to pH 7 using NaOH

Component	Final Concentration	Quantity
Sodium Dihydrogen Phosphate di-hydrate	50mM	7.8g/L
NaCl	100mM	5.84g/L

SP Equilibration Buffer -*adjusted to pH 5 using 100% glacial acetic acid*

Component	Final Concentration	Quantity
Sodium Acetate Trihydrate	50mM	6.8g/L

SP Washing Buffer –*adjusted to pH 5 using 100% glacial acetic acid*

Component	Final Concentration	Quantity
Sodium Acetate Trihydrate	50mM	6.8g/L
NaCl	100mM	5.84g/L

TBS (Tris Buffered Saline) and T-TBS (Tris Buffered Saline – Tween) -*0.5mL of Tween 20 was added per litre of TBS in order to make T-TBS*

Component	Final Concentration	Quantity
Tris Base	50mM	6.05g/L
Potassium Chloride	2.6mM	0.2g/L
Sodium Chloride	137mM	8g/L

Transferrin buffer –*adjusted to pH 8 using 11.6M HCl*

Component	Final Concentration	Quantity
Tris Base	20mM	2.422g/L
$NaHCO_3$	20mM	1.68g/L

Urea Gel TBE Buffer (x20) - *Autoclaved then adjusted to pH 8.4 using 11.6M HCl*

Component	Final Concentration	Quantity
Tris Base	2M	242.2g/L
Boric Acid	0.2M	12.37g/L
EDTA	32mM	11.91g/L

Washing Buffer 1 *–adjusted to pH ~3 with 10M NaOH*

Component	Final Concentration	Quantity
Sodium Acetate	100mM	8.203g/L
NaCl	500mM	29.22g/L

Washing Buffer 2 *–adjusted to pH 8 using 11.6M HCl*

Component	Final Concentration	Quantity
Tris Base	100mM	12.11g/L
NaCl	500mM	29.22g/L

2.1.6. Enzymes

Restriction endonucleases were supplied by Promega Corporation, New England Biolabs, or Roche Applied Sciences. T4 DNA Ligase, T4 Polynucleotide Kinase, Calf Intestinal Alkaline Phosphatase, Taq DNA Polymerase and Pfu DNA Polymerase were all supplied by Promega Corporation. The optimum conditions required for these enzymatic reactions, including choice of buffer, amount of enzyme, and temperature, were obtained from the manufacturers' protocols.

2.1.7. Vectors

Physical maps of the vectors used in these studies are shown in Appendix II.2. A brief description of these vectors' main features are described in this section, but further information can be obtained by visiting the supplier's website.

pALTER-Ex1:

The pALTER-Ex1 vector is a phagemid which is part of the Altered Sites II *in vitro* systems (Promega) that uses a straightforward procedure to generate and select

oligonucleotide-directed mutants. The system uses antibiotic selection as a means to obtain a high frequency of mutants. This vector contains genes for ampicillin and tetracycline resistance, but the ampicillin resistance gene has been inactivated. An oligonucleotide is provided which restores ampicillin resistance to the mutant strand during the mutagenesis reaction, by annealing to the single-stranded DNA (ssDNA) template at the same time as the mutagenic oligonucleotide, and subsequent synthesis and ligation of the mutant strand links the two. As well as mutagenesis, this vector also allows expression of the insert *in vivo* and *in vitro* using a novel polylinker. Transcription of the insert in controlled by opposing T7 and SP6 promoters. A *tac* promoter is also provided in the same direction as the SP6 promoter. To prevent transcription through the tetracycline gene, a T7 transcription terminator is located outside of the multiple cloning-site. The *tac* promoter is induced by the addition of IPTG in a *lac Iq* host, and the expression from T7 can be provided by strains which produce T7 RNA polymerase.

pALTER-Max:

The pALTER-Max vector (Promega) is a plasmid designed for use with Promega's Altered Sites® Mammalian Mutagenesis Systems. It contains the human cytomegalovirus (CMV) immediate-early enhancer/promoter region for strong, constitutive expression of cloned DNA inserts in a variety of mammalian cell types. Mutagenesis using this vector works on the same principal as the pALTER-E1 vector, except that the gene encoding for tetracycline resistance has been substituted with a gene encoding for chloramphenicol resistance. The gene for ampicillin resistance in the pALTER-Max vector has also been deactivated.

pET-12a:

pET12a (Novagen), like all of the other pET vectors, is designed for high-level prokaryotic expression of proteins using a T7 promoter. Expression of proteins using a pET vector is therefore reliant on strains that possess the T7 RNA polymerase. This vector carries an N-terminal *ompT* sequence for potential periplasmic export of the expressed protein. Export into the periplasm is known to provide a more favourable environment for folding and disulphide-bridge formation, therefore increasing the chance of producing active, soluble proteins.

pET-21d:

The pET-21d (Novagen) vector carries an N-terminal T7•Tag® sequence and an optional C-terminal His•Tag® sequence for purification of the target protein.

pDB2514:

pDB2514 (Delta Biotechnology Ltd) is a basic vector containing a pUC18 region for replication in *E. coli,* and a gene encoding apramycin resistance and the human serum transferrin (N413Q, N611Q) cDNA from the translation initiation methionine codon (ATG) to an internal *NheI* site.

pDB2529:

The pDB2529 (Delta Biotechnology Ltd) vector contains the yeast PRB1 promoter (Sleep *et al* 1991) for nitrogen regulated protein expression in yeast, the yeast *ADH1* transcription terminator (Sleep *et al* 1991) and a leader sequence based upon a fusion of the first 19 amino acids of the human serum albumin leader sequence and the last 5 amino acids of the *MFα-1* prepro leader sequence from *S.cerevisiae* for secretion

of protein out of the cell (Sleep *et al* 1990). The vector also carries a gene that encodes for ampicillin resistance, a number of restriction sites required for cloning and a human transferrin cDNA (N413Q, N611Q).

pDB2536:

The pDB2536 (Delta Biotechnology Ltd) vector comprises of the transferrin (N413Q, N611Q) *NotI* expression cassette from pDB2529 cloned into the unique *NotI* site of pSAC35 (Sleep *et al* 1991).

2.1.8. Plasmid DNA Purification and DNA Cleanup Kits

All kits used in these studies were supplied by Qiagen Ltd. These kits included: HiSpeed Plasmid Midi, Spin Mini Prep, 96 Turbo MiniPreps, Mini Elute Gel Extraction and Mini Elute Reaction Cleanup. The only exception was during the cloning of the yeast expression system when a BIO101 Gene Clean III gel extraction kit was used. These kits were used in accordance to the suppliers' protocol.

2.1.9. Loading Buffers

All loading buffers were made up to 100mL with MilliQ water, then filtered (0.2µ).

DNA Agarose-Gel Electrophoresis (5x)

Component	Final Concentration	Quantity
Bromophenol Blue	0.25% (w/v)	0.25g
Sucrose	40% (w/v)	40g

Native-PAGE (5x) (10ml)

Component	Final Concentration	Quantity
1M Tris-Cl pH 6.8	312.5mM	31mL
Glycerol	50% (v/v)	50mL
Bromophenol Blue	0.05% (w/v)	50mg

SDS-PAGE Loading Buffer (2x)

Component	Final Concentration	Quantity
1M Tris-Cl pH 6.8	100mM	10mL
Glycerol	20%(v/v)	20mL
Sodium Dodecyl Sulphate	4.4% (w/v)	4400mg
Bromophenol Blue	0.02% (w/v)	20mg
2-Mercaptoethanol *(For reduced SDS-PAGE only)*	2% (v/v)	2mL

Urea-Gel Loading Buffer (2x)

Component	Final Concentration	Quantity
Glycerol	10% (v/v)	10mL
Bromophenol Blue	0.01% (w/v)	10mg
Urea Gel TBE buffer x 20 *(see 2.1.5. Buffers)*	1 x	5mL

2.1.10. Staining Solutions

PAGE Stain

Component	Quantity
Water	45% (v/v)
Methanol	45% (v/v)
Glacial Acetic Acid	10% (v/v)
Coomasie Blue, or Brillian Blue R	0.25% (w/v)

PAGE Destain

Component	Quantity
Water	80% (v/v)
Glacial Acetic Acid	10% (v/v)
Methanol	10% (v/v)

Ethidium Bromide

The addition of 0.05% (w/v) ethidium bromide to gels enabled DNA to be visualised under UV light

2.1.11. Computer Software

The different software packages used in these studies are listed below, along with a brief description of each one:

-BioEdit (5.0.9): Sequence alignment editor and sequence analyser.

-ChromasPro(1.22): DNA sequence chromatogram analyser

-Deepview/Swiss-PDB viewer: PDB graphical interface and analysis package

-SciEd Central Clone Manager 6: Cloning simulator, enzymes operations and genetic map drawing package. (A gift from Delta Biotechnology Ltd)

2.2. DNA Cloning and Analysis Methods

2.2.1. Glycerol-Stocks

To 700µL of fresh culture and 300µL of sterile 50% (v/v) glycerol was mixed gently in a sterile 1.5mL NUNC cryotubes tube.

2.2.2. DNA Electrophoresis

Agarose gel electrophoresis was performed using a 1% (w/v) agarose gel throughout these studies. The gel was made by the addition of 1g of electrophoresis-grade agarose (Sigma) to 100mL of 1x solution of TBE buffer (section 2.1.5) and then melted in a microwave at full power for 2 minutes. The solution was then allowed to cool to 60°C before the addition of ethidium bromide to a final concentration of 50ng/ml. This agarose was then poured into plastic formers and allowed to set at room temperature. Prior to loading, DNA samples were mixed with 5 x DNA loading buffer (section 2.1.9). Gels were electrophoresed in 1x TBE buffer at 100V constant volts until the dye had reached the end of the gel. Following electrophoresis, gels were examined by transillumination with UV light and photographed on a SynGene GelDoc System.

2.2.3. Ligation

Ligation of DNA was achieved using T4 DNA ligase from Promega. When ligating DNA together a number of different molar ratios of vector and insert were used.

These ratios included 1:1, 1:3 and 3:1 (vector:insert). The reactions were either incubated overnight at 4°C, 4-18 hours at 15°C, or at room temperature for 3 hours.

2.2.4. Transformation

2.2.4.1. Bacterial

Competent *E.coli* cells were removed from -80°C storage and thawed on ice. To 50μL of these cells, 20μL of chilled ligation mixture was added before being incubated on ice for 30 minutes. The cells were then heat shocked at 42°C for 90 seconds prior to the addition of 0.7mL of LB and then incubation at 37°C for 45 minutes. Transformants were selected by growing the cells on LB agar plates overnight at 37°C with the appropriate selective marker.

2.2.4.2. Yeast

Yeast cells grown on YEPD plates were resuspended in a microcentrifuge tube containing 0.5mL of transformation buffer (100mM lithium acetate, 10mM Tris-HCl, pH 7.5, 1mM EDTA pH 8.0). The cells were microcentrifuged for 5 seconds and the supernatant removed, leaving 50-100μL of buffer in the tube. A 10μL aliquot of salmon testes DNA (10mg/ml) and ~1μg of plasmid DNA was added to the cells before they were vortexed for 10 seconds. 600μL of Plate buffer (40% PEG, 100mM lithium acetate, 10mM Tris-HCl, pH 7.5, 1mM EDTA pH 8.0) was then added. The mixture was vortexed again before being incubated at room temperature for 15 minutes. An optional heat shock step for 15 minutes at 42°C can be applied here. Following centrifugation for 3 seconds and the removal of the supernatant, the cells were resuspended in 300μL of sterile water. The cells were then plated on BMM-D media and incubated at 30°C for 2-3 days, or until colonies appeared.

2.2.5. PCR DNA Amplification

PCR amplification of both hLf and hSTf was accomplished in 50μL reaction mixtures containing 1x DNA polymerase buffer with 1.5mM $MgSO_4$, of each of the four dNTPs (200μM), of each primer(1μM), <0.5μg DNA template, 1.25units DNA polymerase and MilliQ water. Both Pfu and Taq DNA polymerase were used in these studies. The choice of DNA polymerase and primers is discussed in Chapter 4. Table 2.3 outlines the thermal cycling conditions used.

Table 2.3. Thermal cycling conditions for hLf and hSTf amplifitcation

Step	Temperature	Time	Number of Cycles
1. Initial Denaturation	95°C	2 minutes	1
2. Denaturation	95°C	1 minutes	30
3. Annealing	58°C	40 seconds	
4. Extension	74°C	2 minutes per 1kb of DNA template	
5. Final Extension	74°C	5 minute	1
6. Soak	4°C	Indefinite	1

2.2.6. UV determination of DNA concentration

DNA concentration of samples was determined by measuring their optical density at 260nm using a Shimadzu UV-1201 UV-VIS spectrophotometer and a 10mm pathlength curvette. Samples were diluted 50-fold using MilliQ water.

1.0 OD at 260nm = 50μg/mL *(Promega)*

2.2.7. Mutagenesis

Mutagenesis of hSTf cDNA was achieved using a Stratagene Quikchange® Site-Directed Mutagenesis Kit. The two oligonucleotides required for mutagenesis were designed using Sci Ed Central Clone Manager 6 software and synthesised by Sigma-

Genosys. Following the manufacturers' protocol, a 50μL reaction mix comprising 5μL of QuikChange Reaction buffer (x10 concentration), 10ng of template plasmid, 125ng of both mutagenesis oligonucleotides, 1μL of dNTP Mix, 1μL of *PfuTurbo®* DNA polymerase, and MilliQ water was made. This reaction mix was then incubated in a DNA Engine DYAD™ Peltier thermocycler using the program described in Table 2.4. Following incubation, 1μL of *Dpn*1 was added and the mixture incubated at 37°C for 1 hour.

Table 2.4. The thermocycler program for QuikChange site-directed mutagenesis of pDB2514.

Step	Temperature	Time	Number of Cycles
1.Initial Denaturation	95°C	30 seconds	1
2. Denaturation	95°C	30 seconds	16
3. Annealing	55°C	1 minute	
4. Extension	68°C	5 minutes	
6. Hold	4°C	Indefinite	1

2.2.8. DNA Sequencing

DNA sequencing was either carried out using a Perkin Elmer ABI Prism 310 Genetic Analyser. Each 20μL PCR reaction comprises 500ng of DNA template, 0.8pmol/μL of primer, 8μL of ABI Prism Big Dye V3 Terminator and MilliQ water. The thermocycle program is shown in Table 2.5.

Table 2.5. Thermal cycling conditions for ABI Prism Big Dye Sequencing

Step	Temperature	Time	Number of Cycles
1. Denaturation	96°C	10 seconds	25
2. Annealing	50°C	5 seconds	
3. Extension	60°C	4 minutes	
4. Soak	4°C	Indefinite	1

The reaction from the PCR tube was then transferred into a 1.5mL eppendorf. To this reaction was added 2μL of 3M NaOAc, pH 4.6 and 50μL of 95% EtOH. The sample was mixed and then left to stand for 15 minutes at RT, before being centrifuged for 20 minutes at maximum speed. Following centrifugation, the supernatant was immediately removed and 750μL of 70% EtOH added to the pellet. The sample was vortexed briefly and then centrifuged for 5 minutes, at maximum speed. Once the supernatant had been removed, the pellet was allowed to dry rapidly on a 95°C heating block, to remove any remaining ethanol. The pellet was resuspended in 12μL of Template Suppression Reagent (Perkin Elmer), vortexed briefly, denatured at 95°C for 2 minutes, and then placed immediately on ice. Samples were centrifuged briefly before being loaded onto the ABI Prism 310 Genetic Analyser.

2.3. Protein Purification and Preparation Methods

2.3.1. Ion-Exchange Chromatography

Two ion-exchange chromatography columns were used in these studies to purify rhSTf. The first was a 232mL SP-FF sepharose (Amersham Biosciences) column and the second was a 94mL DEAE-FF sepharose (Amersham Biosciences) column. The methodology described below was created by Delta Biotechnology Ltd. Both columns were connected to a Pharmacia Biotech Pump P-1 and a Single Path Monitor UV-1 Optical / Monitor Unit measuring at 280nm, with an output to a Pharmacia Chart Recorder. Columns were stored in 0.5M NaOH when not in use.

2.3.1.1. SP-FF Sepharose

Once the column had been equilibrated with SP Equilibration Buffer at 0.5 cv/min for 8 minutes, or until the pH of the eluate was approximately 5, the sample was

loaded on to the column at the same rate. The column was then washed at 0.5cv/min with SP Equilibration Buffer for 10 minutes, SP Washing Buffer for 16 minutes and again with SP Equilibration Buffer for 4 minutes. The last wash was extended if the absorbance at 280nm had not yet reached zero. Using a reduced flow rate of 0.25cv/min, rHSTf was then eluted with SP Elution Buffer, collecting the flow through only when the UV spectrophotometer recorded a peak at 280nm. Once the absorbance had reached baseline, the column was washed at 0.25cv/min with 1M NaCl for 12 minutes, then with 0.5M NaOH for a further 12 minutes.

2.3.1.2. DEAE-FF Sepharose

The flow rate in which this column was run was set at a constant 0.5cv/min. Prior to loading, the column was equilibrated with DEAE Equilibration Buffer 1 for 24 minutes and then with DEAE Equilibration Buffer 2 for a further 6 minutes. Once the sample was loaded, the column was washed with DEAE Equilibration Buffer 2 for 10 minutes, or until the absorbance at 280nm returned to baseline. The rhSTf was then eluted with DEAE Elution Buffer, starting collection of the eluate when the UV spectrophotometer recorded a peak at 280nm. Following elution, the column was washed with 1M NaCl for 6 minutes and then with 0.5M NaOH for 4 minutes.

2.3.2. Immunoaffinity Chromatography

CNBr-activated Sepharose™ 4B was purchased from Amersham Biosciences. Each affinity column was made using 15g of freeze-dried medium. This amounted to approximately 50mL once the beads had been swollen, by washing in 1mM HCl.

Ligand coupling was achieved by dissolving the ligand in Coupling Buffer and then adding the required amount of medium. The mixture was incubated at room temperature for one hour with gentle mixing, not stirring. Following this, the medium was washed clean of any unbound ligands using 5cv of Coupling Buffer and then incubated at room temperature for 2 hours in Immunoaffinity Buffer to ensure that any remaining active groups are blocked. The final washing step comprised three cycles of alternating pH. Each cycle consisted of a wash with Washing Buffer 1 followed by a wash with Washing Buffer 2. A total of 5 column volumes of each buffer was used in the final washing step. The ligand-coupled medium was then packed into a Pharmacia XK26 column, connected to a Pharmacia Biotech Pump P-1 and a Single Path Monitor UV-1 Optical / Monitor Unit with an output to a Pharmacia Chart Recorder.

The affinity column was equilibrated using 8 column volumes of Immunoaffinity Buffer Following this, the sample, which had been filtered through a 0.2μm filter and diluted 1:1 with 2x Equilibration Buffer, was loaded onto the column at 4ml.min^{-1}. The column was then washed with 10cv of Immunoaffinity Buffer, or until the OD at 280nm reached baseline, before eluting the protein using 1M ammonia.

2.3.3. Fast Performance Liquid Chromatography

A Pharmacia Fine Chemical FPLC system was used in these studies. This setup consisted of a Gradient Programmer GP-250, two P-500 pumps (pump A and B), a Rec-2 Chart Recorder, UV-M monitor with optional 214nm / 280nm filters, a mixer, V-7 Valve, a 2mL loading-loop and a RediFrac fraction collector. Before use, the two pumps were washed with their respective buffers and the loading-loop washed

with MilliQ water. All samples and buffers were filtered through a 0.2μm filter before use.

2.3.4. Protein Concentration and Dialysis

Protein samples were concentrated and dialysed using an Amicon Ultra-15 (30,0000MWCO) centrifugal filter units at 4000g for 15 minutes in a swingout rotor. Centrifugation was repeated for samples larger than 15mL until the required concentration and volume was achieved. Dialysis was achieved by repeatedly resuspending and concentrating the sample in the specific buffer.

2.3.5. Creation of Apo- and Diferric Transferrin

Apo-hSTf preparations were made by dialysing hSTf against Citric Acid Buffer (section 2.1.5) to release the iron from the protein. These samples were then left to incubate at room temperature on a roller for 2 hours, before further dialysis against the same buffer to remove any free iron. Apo-hSTf was then dialysed into Transferrin Buffer (section 2.1.5) before use. Diferric preparations of protein were made by dialysing hSTf against Transferrin Buffer and then saturating by addition of two-fold excess of FeNTA (section 2.1.5) in excess. Dialysing the samples further into the same buffer washed unbound FeNTA away.

2.3.6. Desalting and Freeze-drying

Purified proteins were desalted into 50mM NH_4HCO_3 using a G-25 Sephadex (Amersham Biosciences) column. The protein solution was then transferred into a round bottom Quick-fit vacuum flask and frozen evenly onto its inner surface using liquid nitrogen. The flask was then connected to the Modulyo freeze-dryer

(Edwards), previously cooled to -45°C and depressurised to 0.08mBar, and left overnight, or until the solvent had been removed.

2.4. Protein Analysis Methods

2.4.1. Circular Dichroism (CD) Spectroscopy

CD spectra for protein samples were acquired on the Jasco J720 spectropolarimeter in the 600-230nm (1cm cell path length) and 260-185nm (0.02cm cell path length) regions. Protein samples were made to 0.5mg/mL in 0.02M $NaHCO_3$ and filtered through a 0.2μm syringe filter. Protein concentrations were calculated after light scattering correction had been applied. All spectra were corrected for solvent baseline and, where appropriate, normalised for concentration and path length (mean molecular weight = 113). Secondary structure analysis of the proteins was calculated using the principle component analysis method (Malik, 1997).

2.4.2. Cell Proliferation Assay

Cellular proliferation was determined by quantification of 5-bromo-2'-deoxy-uridine (BrdU) integrated into cellular DNA. TRVb-1 (McGraw *et al*, 1987) (kindly provided by Dr. H. Kawabata, Kanazawa Medical University Uchinada-machi, Ishikawa-ken Japan) or HepG2 cells were seeded in 6-well tissue culture plates at a density of 5×10^4 cells per well and grown to 50% confluence in F12HAM (Invitrogen) and RPMI160 (Invitrogen), respectively supplemented with 10% foetal bovine serum (FBS) 2mM Glutamine and Penicillin/Streptomycin. Cells were then synchronised by cultivation in the absence of FBS for 24 hours. Immediately after addition of transferrin, to a final concentration of 2mg/ml, or 5% FBS to the cells, BrdU (Roche Applied Sciences) at a final concentration of 10μM was added and the

cells were grown for a further 16 hours. Integration of BrdU into cellular DNA was then quantified immuno-metrically using the BrdU detection Kit III (Roche Applied Sciences) according to the manufacturer's directions.

2.4.3. Dot Blot Analysis

Dot blots were used in these studies to determine levels of protein expression and to visualise protein-protein interactions. Nitrocellulose Membrane from BioRad (Trans-Blot Transfer Medium) was cut to size and marked with a pencil for orientation and location of drops. For each drop, 4μL of protein sample was pipetted onto the membrane and allowed to dry at room temperature. The membrane was then washed in Nitrocellulose Blocking Buffer for 1 hour on a rocking platform to block the remaining membrane surface. Excess Nitrocellulose Blocking Buffer was removed by 3 x 5 min washes with PBS-T before the membrane was incubated for 2 hours in 10mL of PBS containing 1μg/mL FITC labelled IgG specific for the protein of interest. A final wash with PBS to remove any unbound FITC-IgG was carried out before visualisation of the bound FITC-IgG under UV light.

2.4.4. Dynamic Light Scattering

Protein samples were prepared at 0.1mg/mL in Transferrin Buffer and filtered through a 0.2μm syringe filter prior to analysis. Samples (100μl) were analysed in a sealed quartz curvette using a DynaPro99 and Temperature Controlled Microsampler (Protein Solutions Inc). Using the DynaPro Dynamic Ligtht Scattering instrument control software, each sample was initially cooled to 5°C and then held at this temperature for 2 minutes. The machine then took 20 photon count recordings before heating the sample up to the next temperature and holding it a further 2 minutes.

After 5°C, the machine ramped the sample up to 15°C and then 5°C increments up to 95°C. Results were analysed automatically using the Dynamic (V5.26.37) Software package.

2.4.5. Electron Paramagnetic Resonance (EPR) Spectroscopy

EPR spectra were obtained using a Bruker ELEXSYS E500 spectrometer. The protein samples (40mg/ml) in 50mM $NaHCO_3$, were frozen in an EPR tube at 77°K in liquid nitrogen. The microwave frequency was maintained at 9.38GHz and the field set at 2499G with modulation amplitude at 5G. The sample was then scanned twice using a microwave power of 10mW over 4000G (400mT) with a receiver gain of $2x10^3$G.

2.4.6. Fluorescein Isothiocyanate (FITC) Protein Labelling

FITC Isomer 1 on Celite from Sigma was used to label immunoglobulin G (IgG) molecules specific for Transferrins. Freeze dried immunoaffinity chromatography purified IgG was resuspended in 0.1M sodium carbonate pH 9 at 2mg/ml. To this solution, FITC-celite resuspended in DMSO was added in 5µL aliquots to the protein solution to a molarity three times that of the protein. The amount of FITC required was multiplied ten times to compensate for the mass of the celite. Following overnight incubation at 4°C on a roller machine, ammonium chloride was added to a final concentration of 50mM before incubation for a further 2 hours at 4°C on a roller machine. All incubations were carried out in the dark. Xylene Cyanol to 0.1% (w/v) and glycerol to 5% (v/v) was then added before the mixture was span at 5000*g* for 30 min to pellet the Celite. Unbound FITC was separated from the conjugate by gel

filtration of the supernatant on a G-25 Sephadex column wrapped in aluminium foil to block out the light. Samples were stored in the dark at 4°C.

Determination of Fluorescein: IgG Molar Ration

$$\text{Molar}(f/p) = \frac{2.77 \times A_{495}}{A_{280} - (0.35 \times A_{495})}$$

Determination of FITC - IgG concentration

$$\text{IgG}(mg/ml) = \frac{A_{280} - (0.35 \times A_{495})}{1.4}$$

Information taken from FITC product sheet: Sigma-Aldrich

2.4.7. High Performance Liquid Chromatography (HPLC)

To work out protein concentrations and to determine relative protein purity, samples were separated by size-exclusion chromatography using a G3000SWXL (Tosoh Biosciences) column and a Shimadzu LC2010 HPLC. A Shimadzu-SIL-HTA Autosampler was used to load samples. The HPLC machine was calibrated using hSTf (Calbiochem) as an internal reference to allow the elution time and protein concentration to be determined. Protein concentration was determined by measuring the area of the curve created by plotting absorption at 280nm against time.

2.4.8. Immunodiffusion

A thin layer of 1% (w/v) Agar (VWR) and 20mM Tris molten solution was gently poured onto a clean glass microscope slide, and left to set at room temperature. Once set, a template was used to accurately cut five holes in the agar in the same pattern as on the fifth-side of a dice. In to each hole, 10μL of sample (1mg/mL) was pipetted

and the slide sealed, along with a small piece of moist filter paper, in a petri dish with parafilm. The slide was then incubated at 4°C for 24-48 hours.

2.4.9. Isothermal Titration Calorimetry (ITC)

All protein samples to be analysed were prepared in ITC sample buffer (section 2.1.5) at 2mg/mL (~26μM) and then filtered using a 0.2μm syringe filter. 0.1M FeNTA (section 2.1.5) was diluted into ITC sample buffer to 500μM and then filtered using a 0.2μm syringe filter. Both protein samples and the FeNTA were degassed and heated to 25°C using a ThermoVac Degasser/thermostat (MicroCal).

Experiments were performed using the VP-ITC (MicroCal, USA) in the buffer conditions given above. Before titration, the machine was allowed to heat both cells to exactly 25°C. For each titration 1.463mL of protein, in the cell, was titrated with 19 injections of 500μM FeNTA. The volume of each injection was 15μL, except for the first injection, which was 2μL. Injections were continued beyond saturation levels to allow for determination of heats of ligand dilution. The resulting data were fit to a double-site binding isotherm using the ORIGIN software supplied with the calorimeter. All thermodynamic data was generated automatically by the software.

2.4.10. Native-PAGE

Discontinuous non-denaturing (native) PAGE was carried out using the BioRad Mini-Protean II system. The resolving gel (8%) comprised 2.5mL of 1.5M Tris (pH 8 with HCl), 2.66mL of Acrylamide stock solution (30% acrylamide, 0.8% Bis-acrylamide) and 50μL of 10% ammonium persulfate, made up to 10mL with MilliQ water. The stacking gel (5%) comprised 0.67mL of Acrylamide stock solution, 1mL

of 0.5M Tris (pH 6.8 with HCl) and 30μL of 10% ammonium persulfate, made up to 4mL with MilliQ water. 5μL of TEMED was added to each solution before gels were poured. Samples were prepared by addition of an equal volume of native-PAGE loading buffer (section 2.1.9) prior to loading on the gel. The gels were run at 20mA, constant current, in a BioRad tank with 1x Native-PAGE Running buffer (section 2.1.9) until the loading dye reached the bottom of the gel. As with SDS-PAGE, bands were visualised by submerging the gels in PAGE stain (section 2.1.10) for 15 minutes on a rocking platform and then destained by replacing the stain with PAGE Destain solution (section 2.1.10) and rocking the gel for a further 45 minutes, or until the background was clear.

2.4.11. Protein Crystallisation

Purified protein was concentrated to the relevant concentration using the method described in section 2.3.4 and filtered using a 0.2μm syringe filter. The precipitant solutions were prepared to the relevant concentration in MilliQ H_20 and passed through a 0.2μm syringe filter following adjustment to the appropriate pH.

2.4.11.1. Hanging Drop (Vapour diffusion) Method

Hanging drops were set up using 24-well Limbro plates (Hampton Research). Aliquots (1mL) of precipitant buffer were added to each of the 24 wells. Grease was then applied to the top of each well using an air pump. A drop (3μl) of protein sample was pipetted onto a siliconised cover slip and immediately mixed with 3μL of the relevant precipitant buffer. Care was taken not to introduce any air bubbles into the drop. The cover slip was then inverted and placed over the relevant well,

ensuring that an airtight seal had formed around the grease. The trays were then stored at either 277K or 293K until crystals appeared.

2.4.11.2. Microbatch Method

Microbatch assays were set up using 74-well Terasaki Plates (Hampton Research). Mineral oil was poured onto the plate to cover all wells by at least 2mm. For each crystallisation trial, the protein sample of interest (1µl) was mixed with the relevant crystallisation agent (1µl) and then dropped through the oil into the relevant well using an automatic pipette. Microbatch plates were then at either 277K or 293K until crystals appeared.

2.4.12. SDS-PAGE

SDS-PAGE analysis used a discontinuous method adapted from Laemmli (1970). 8% acrylamide resolving gels and 5% stacking gels were used with the BioRad Mini-Protean II system. Samples were prepared by adding an equal volume of SDS-PAGE loading buffer (section 2.1.9) followed by boiling for eight minutes prior to loading on the gel. The gels were run at 20mA, constant current, in a BioRad tank with 1x SDS running buffer (section 2.1.5) until the loading dye reached the bottom of the gel. To visualise the bands, the gels were submerged in PAGE stain (section 2.1.10) for 15 minutes on a rocking platform and then destained by replacing the stain with PAGE destain solution (section 2.1.10) and rocking the gel for a further 45 minutes, or until the background was clear.

2.4.13. SPOTs Peptide Library Mapping

Before use, each of the SPOTs membranes (Sigma Genosys) was allowed to warm to room temperature and then rinsed in methanol. Each membrane was then sealed in a polypropylene box and rocked for 10 minutes in 50mL of TBS. This washing procedure was repeated twice.. The membrane was then incubated at room temperature overnight, in 50mL of 1x SPOTs Blocking Buffer (Sigma Genosys), before being washed for 10 minutes in 50mL of T-TBS. The membrane was then left to incubate for 4 hours on a rocking platform in 50mL of SPOTs blocking buffer containing Fluorescein (FITC)-conjugated Chromopure hSTf (Jackson ImmunoResearch) at a final concentration of 5µg/mL. Unbound FITC-hSTf was removed by washing the membrane three times for 10minutes in 50mL of T-TBS. The binding of FITC-hSTf to the SPOTs membrane was visualised using a UV SynGene GelDoc System and quantified using SynGene Genetools. Membranes were cleaned following the manufacturers' guidelines.

2.4.14. 6M Urea PAGE

Polyacrylamide gel electrophoresis in 6M urea was carried out using the method described in Evans *et al,* 1978 and Evans *et al,* 1982. Table 2.6 shows the composition of a 6M urea gel. This solution was made up to 40mL with MilliQ water and the components dissolved by vortexing. Prior to pouring the gel between a two-plate setup, 60µL of TEMED was added and the solution inverted quickly several times to mix. All components, except for the Urea and TBE, were supplied by National Diagnostics.

Table 2.6. The composition of a 6M urea gel.

Component	Molarity / Concentration	Quantity
Urea	6M	14.4g
Bis-AcrylaGel™	0.32% (w/v)	6.4ml
Protogel	6.5% (w/v)	8.7ml
Urea Gel TBE Buffer (20x)	1 x	2ml
Ammonium Persulfate	0.1% (w/v)	40mg

Once the gel had set, the comb removed and the wells washed with water, 50µL samples (25µL of 0.65mg/mL protein sample and 25µL of Urea Gel loading buffer) were loaded onto the gel. The gel was then run at 100V (constant) for 16hrs, before being removed from the plates and stained and destained in an identical manner to the SDS gels.

2.4.15. UV-Visible Scanning Spectroscopy

Apo-hSTf was prepared at ~10mg/mL in Transferrin Buffer and scanned in a 10mm path length quartz cuvette from 350-600nm using a Perkin Elmer Lambda 2 UV-visible spectrometer. The protein was then titrated with 5µL aliquots of 0.01M FeNTA, repeating the scan between aliquots, until the protein reached saturation. In order to determine protein concentration throughout these studies, the extinction coefficients at 280nm for apo and diferric hSTf as $88200M^{-1}cm^{-1}$ and $114000M^{-1}cm^{-1}$, respectively (Aisen *et al*, 1980), and at 470nm for diferric hSTf, $4620M^{-1}cm^{-1}$ (Bates and Wernicke, 1971) was used in the following equation:

$$c = \left(\frac{A_{280}}{\varepsilon}\right) M_w B$$

To determine the absorbance of a sample at a known protein concentration, the following equation was used:

$$A = \varepsilon B\left(\frac{c}{Mw}\right)$$

Where:

ε	extinction coefficient
B	path length of quartz cuvette (1cm)
C	protein concentration (mg/ml)
Mw	Molecular Weight (hSTf =79,550Da)
A_{280}	Absorbance at 280nm

3. Computational Analysis of Transferrins

This chapter focuses on the *'in silico'* characterisation of hSTf through sequence analysis and homology modelling. The information from this is then to be used as a tool to help in the interpretation of the *in vitro* data obtained from both the recombinant and native proteins, which are described in the later chapters. The compilation of theoretical predictive modelling with the experimental results will take us a step forward towards a complete understanding of the mechanisms of iron-uptake/release of transferrin.

3.1. Sequence Analysis and Homology Modelling

Even though much is known about hSTf, no crystallographic structure of the full-length hSTf molecule is available yet. This prevents us from answering a number of important questions concerning the mechanism of iron binding and release and also the mechanism of delivering iron to the cells. Therefore, in order to visualise this molecule and gain a better understanding of it, the initial aim of this chapter is to create a reliable computer 3D model of the protein. Such a model will be an essential tool to be used for the interpretation of the results obtained in structural and functional experiments carried out later on in these studies on native and recombinant hSTf, with both wild-type and mutant variants.

3.1.1. Primary Sequence Analysis

As described in Chapter 1, hSTf is a monomeric bilobal glycoprotein consisting of 679 amino acids with a 19 amino acid leader that is processed off in the mature protein (Appendix I). This protein can be divided into two lobes, the N-lobe and the

C-lobe, which spans form residues 1-336 and from residues 337-679 respectively (MacGillivray *et al* 1983), and are connected by a bridge region from residues 331 to 339 (Ugo Testa, 2002).

3.1.1.1. *Blast* Search

The first step in the creation of a model is to identify solved structures of proteins that are homologous to hSTf. From these results, the molecule with the highest degree of homology can then be used as a template onto which the model can be based. In order to find solved structures of proteins that share a high degree of sequence homology with hSTf, its sequence was extracted from the NCBI website (Accession number: NP_001054) and a similarity search was carried out using BLASTp (NCBI website) over the PDB database (Brookhaven Protein Data Bank).

The results of this search (Figure 3.1) identified rabbit serum transferrin (rSTf) with the highest identity score of 78%, followed by porcine serum transferrin (pSTf) with an identity score of 70%. Below pSTf, the next four closest molecules are all lactoferrins. It is noteworthy a lower identity score was obtained with human lactoferrin (hLf) than with the other lactoferrins from different organisms.

A phylogram of these six proteins, hSTf and eighteen other transferrins was created using ClustalW (Phylip) (www.ebi.ac.uk/clustalw) and edited using TreeviewX (Figure 3.2). From the results obtained with the phylogram, which illustrates the evolutionary relationship of these twenty-five proteins, it can be concluded, in agreement with the BLASTp results, that out of these Transferrins, rSTf is the closest relative to hSTf.

i. **1JNFA:** Chain A, **Rabbit Serum Transferrin** At 2.6 A Resolution.
Length = 676
Score =1114 bits (2881), Expect = 0.0
Identities = 532/678(78%), Positives = 595/678(87%), Gaps = 3/678(0%)

ii. **1H76A**: Chain A, The Crystal Structure Of Diferric **Porcine Serum Transferrin**
Length = 696
Score = 986 bits (2548), Expect = 0.0
Identities = 484/686(70%), Positives = 560/686(81%), Gaps = 12/686(1%)

iii. **1F9BA**: Chain A, Melanin Protein Interaction: X-Ray Structure Of The Complex Of **Mare Lactoferrin** With Melanin Monomers
Length = 695
Score = 858 bits (2218), Expect = 0.0
Identities = 432/693(62%), Positives = 527/693(76%), Gaps = 22/693(3%)

iv. **1JW1A**: Chain A, Crystallization And Structure Determination Of **Goat Lactoferrin** At 4.0 Resolution: A New Form Of Packing In Lactoferrins With A High Solvent Content In Crystals
Length = 689
Score = 839 bits (2168), Expect = 0.0
Identities = 421/690(61%), Positives = 510/690(73%), Gaps = 20/690(2%)

v. **1BIY:** Structure Of Diferric **Buffalo Lactoferrin**
Length = 689
Score = 838 bits (2165), Expect = 0.0
Identities = 422/690(61%), Positives = 508/690(73%), Gaps = 20/690(2%)

vi. **1CB6A:** Chain A, Structure Of **Human Apolactoferrin** At 2.0
Length = 691
Score = 834 bits (2155), Expect = 0.0
Identities = 420/692(60%), Positives = 509/692(73%), Gaps = 22/692(3%)

Figure 3.1. The six most similar molecules in the PDB database (Brookhaven Protein Data Bank) to hSTf (NP_001054), determined by BLASTp (NCBI).

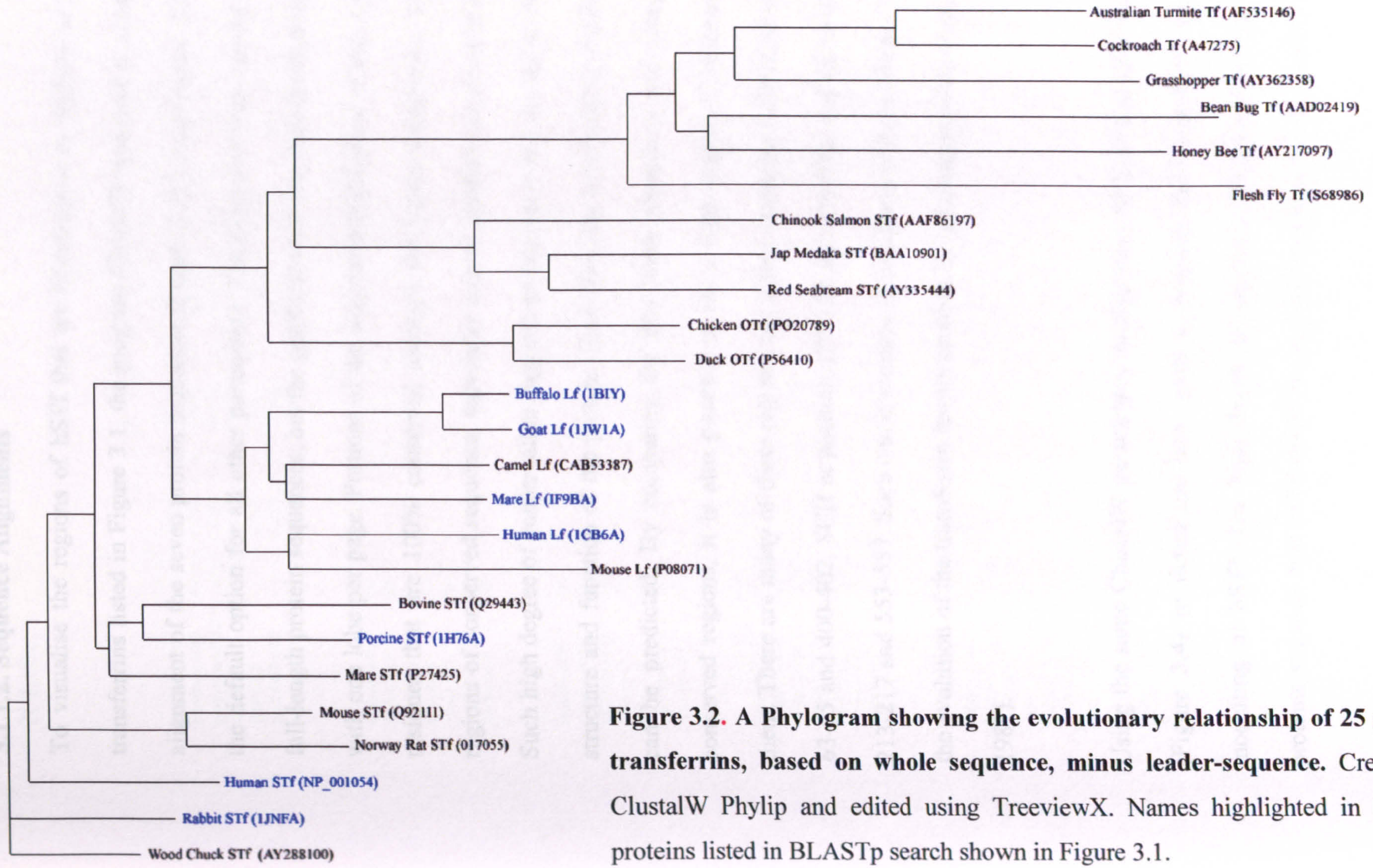


Figure 3.2. A Phylogram showing the evolutionary relationship of 25 sequenced transferrins, based on whole sequence, minus leader-sequence. Created using ClustalW Phylip and edited using TreeviewX. Names highlighted in blue relate proteins listed in BLASTp search shown in Figure 3.1.

3.1.1.2. Sequence Alignments

To visualise the regions of hSTf that are homologous to regions in the other six transferrins listed in Figure 3.1, the program *ClustalW* was used to create a multiple alignment of the seven protein sequences (Figure 3.3), using Gonnet 250 matrix and the default option for all other parameters. The alignment was carried out over the full-length protein sequences, but the graphical view was then divided into each lobe, with one lobe per page. Portions of the sequences highlighted in red, which identify residues that are 100% conserved amongst the seven transferrins, indicate large regions of conserved sequence, especially around residues involved in iron binding. Such high degree of conservation suggests an important role for these residues in the structure and function of the protein. From this, the iron-binding residues for hSTf can be predicted. By comparing the two lobes separately and identifying their conserved regions, it is also possible to see a high degree of homology between them. There are many of these regions, but a few examples include ADA at position 63-65 and 400-402, SCH at position 123-125 and 470-472, and GDVAF at position 213-217 and 553-557. Such a phenomenon is consistent with the theory that suggests the evolution of the transferrin family proteins by gene-duplication events (Park *et al* 1985).

Using the same ClustalW parameters, an alignment of hSTf and rSTf was generated (Figure 3.4) to determine how good a template rSTf would be for homology modelling of hSTf. The high degree of sequence homology between these two proteins suggests that rSTf would be an ideal template for the modelling of hSTf.

N-Lobe

```
                    10        20        30        40        50        60
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
humanSTf        ------VPDK TVRWCAVSEH EATKCQSFRD HMKSVIPSDG PSVACVKKAS YLDCIRAIAA  54
rabbitSTf       ------VTEK TVRWCAVNDH EASKCANFRD SMKKVLPEDG PRIICVKKAS YLDCIKAIAA  54
porcineSTf      ------VAQK TVRWCTISNQ EANKCSSFRE NMSKAVK-NG PLVSCVKKSS YLDCIKAIRD  53
mareLf          LGLCLAAPRK SVRWCTISPA EAAKCAKFQR NMKKVRG--- PSVSCIRKTS SFECIQAIAA  57
goatLf          ------APRK NVRWCAISLP EWSKCYQWQR RMRKLGA--- PSITCVRRTS VLECIRAIAG  51
buffaloLf       ------APRK NVRWCTISQP EWLKCHRWQW RMKKLGA--- PSITCVRRAF VLECIRAITE  51
humanLf         ------GRRR SVQWCAVSQP EATKCFQWQR NMRKVRG--- PPVSCIKRDS PIQCIQAIAE  51

                    70        80        90        100       110       120
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
humanSTf        NEADAVTLDA GLVYDAYLAP NNLKPVVAEF YGSKEDPQTF YYAVAVVKKD SGFQMNQLRG  114
rabbitSTf       HEADAVTLDA GLVHEAGLTP NNLKPVVAEF YGSKENPKTF YYAVALVKKG SNFQLNELQG  114
porcineSTf      KEADAVTLDA GLVFEAGLAP YNLKPVVAEF YGQKDNPQTH YYAVAVVKKG SNFQWNQLQG  113
mareLf          NKADAVTLDG GLVYEAGLHP YKLRPVAAEV YQTRGKPQTR YYAVAVVKKG SGFQLNQLQG  117
goatLf          KNADAVTLDD GMVFEAGRDP YKLRPVAAEI YGTEKSPQTH YYAVAVVKKG SNFKLDQLQG  111
buffaloLf       KKADAVTLDG GMVFEAGLDP YKLRPVAAEI YGTKESPQTH YYAVAVVKKG SNFQLDQLQG  111
humanLf         NRADAVTLDG GFIYEAGLAP YKLRPVAAEV YGTERQPRTH YYAVAVVKKG GSFQLNELQG  111

                    130       140       150       160       170       180
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
humanSTf        KKSCHTGLGR SAGWNIPIGL L--YCDLPEP RKPLEKAVAN FFSGSCAPCA DGTDFPQLCQ  172
rabbitSTf       KKSCHTGLGR SAGWNIPIGL L--LCDLPEP RKPLEKAVAS FFSGSCVPCA DGADFPQLCQ  172
porcineSTf      KRSCHTGLGR SAGWIIPMGL L--YDQLPEP RKPIEKAVAS FFSSSCVPCA DPVNFPKLCQ  171
mareLf          VKSCHTGLGR SAGWNIPIGT LRPYLNWTGP PEPLQKAVAN FFSASCVPCA DGKQYPNLCR  177
goatLf          QKSCHMGLGR SAGWNIPVGI LRPPLSWTES AEPLQGAVAR FFSASCVPCV DGKAYPNLCQ  171
buffaloLf       RNSCHTGLGR SAGWNIPMGI LRPYLSWTES LEPLQGAVAK FFSASCVPCV DRQAYPNLCQ  171
humanLf         LKSCHTGLRR TAGWNVPIGT LRPFLNWTGP PEPIEAAVAR FFSASCVPGA DKGQFPNLCR  171

                    190       200       210       220       230       240
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
humanSTf        LCPG-----C GCSTLNQYFG YSGAFKCLKD GAGDVAFVKH STIFENLANK ADRDQYELLC  227
rabbitSTf       LCPG-----C GCSSVQPYFG YSGAFKCLKD GLGDVAFVKQ ETIFENLPSK DERDQYELLC  227
porcineSTf      QCAGKGAEKC ACSNHEPYFG YAGAFNCLKE DAGDVAFVKH STVLENLPDK ADRDQYELLC  231
mareLf          LCAGTEADKC ACSSQEPYFG YSGAFKCLEN GAGDVAFVKD STVFENLPDE AERDKYELLC  237
goatLf          LCKGVGENKC ACSSQEPYFG YSGAFKCLQD GAGDVAFVKE TTVFENLPEK ADRDQYELLC  231
buffaloLf       LCKGEGENQC ACSPREPYFG YSGAFKCLQD GAGDVAFVKE TTVFENLPEK ADRDQYELLC  231
humanLf         LCAGTGENKC AFSSQEPYFS YSGAFKCLRD GAGDVAFIRE STVFEDLSDE AERDEYELLC  231

                    250       260       270       280       290       300
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
humanSTf        LDNTRKPVDE YKDCHLAQVP SHTVVARSMG GKEDLIWELL NQAQEHFGKD KSKEFQLFSS  287
rabbitSTf       LDNTRKPVDE YEQCHLARVP SHAVVARSVD GKEDLIWELL NQAQEHFGKD KSGDFQLFSS  287
porcineSTf      RDNTRRPVDD YENCYLAQVP SHAVVARSVD GQEDSIWELL NQAQEHFGRD KSPDFQLFSS  291
mareLf          PDNTRKPVDA FKECHLARVP SHAVVARSVD GREDLIWKLL HRAQEEFGRN KSSAFQLFKS  297
goatLf          LNNTRAPVDA FKECHLAQVP SHAVVARSVD GKENLIWELL RKAQEKFGKN KSQRFQLFGS  291
buffaloLf       LNNTRAPVDA FKECHLAQVP SHAVVARSVD GKEDLIWKLL SKAQEKFGKN KSGSFQLFGS  291
humanLf         PDNTRKPVDK FKDCHLARVP SHAVVARSVN GKEDAIWNLL RQAQEKFGKD KSPKFQLFGS  291

                    310       320       330       340       350
                ....|....| ....|....| ....|....| ....|....| ....|....|
humanSTf        PHG-KDLLFK DSAHGFLKVP PRMDAKMYLG YEYVTAIRNL REGTCPEAPT . . . . . 336
rabbitSTf       PHG-KNLLFK DSAYGFFKVP PRMDANLYLG YEYVTAVRNL REGICPDPLQ . . . . . 336
porcineSTf      SHG-KDLLFK DSANGFLKIP SKMDSSLYLG YQYVTALRNL REEISPDSSK . . . . . 340
mareLf          TPEEQDLLFK DSALGFVRIP SQIDSGLYLG ANYLTATQNL RE--TAAEVA . . . . . 345
goatLf          PEGRRDLLFK DSALGFLRIP SKVDSALYLG SRYLTALKNL RE--TAEEVK . . . . . 339
buffaloLf       PPGQRDLLFK DSALGFLRIP SKVDSALYLG SRYLTALKNL RE--TAEEVQ . . . . . 339
humanLf         PSGQKDLLFK DSAIGFSRVP PRIDSGLYLG SGYFTAIQNL RK--SEEEVA . . . . . 339
```

Figure continues on the next page....

C-Lobe

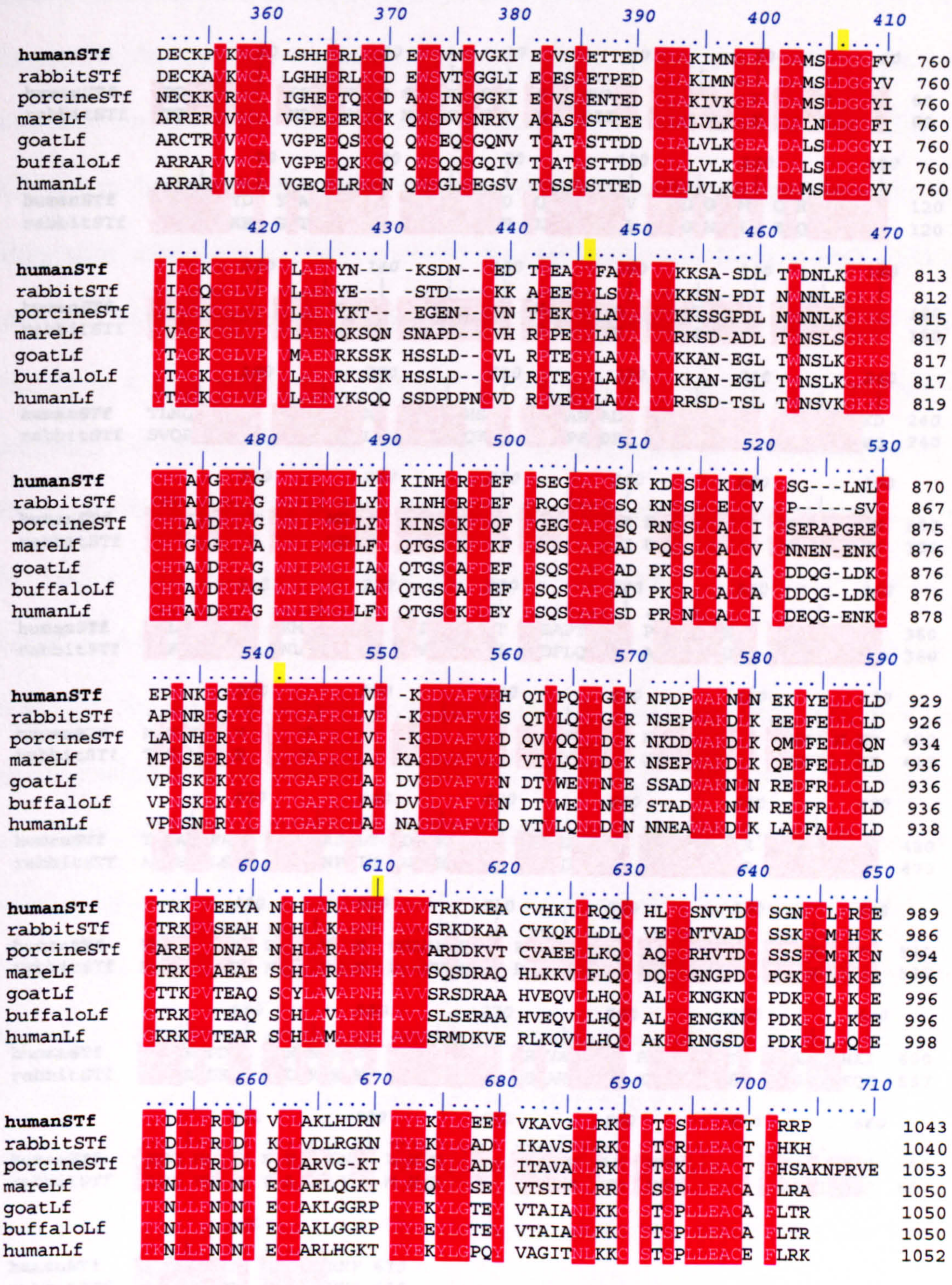


Figure 3.3. Multiple alignment created by ClustalW of hSTf and the six most similar molecules determined following BLASTp. The N- and C-lobes are divided between the two pages. Sequence in red highlights regions of 100% identity. The eight iron-binding residues for all of the seven transferrins are marked above in yellow.

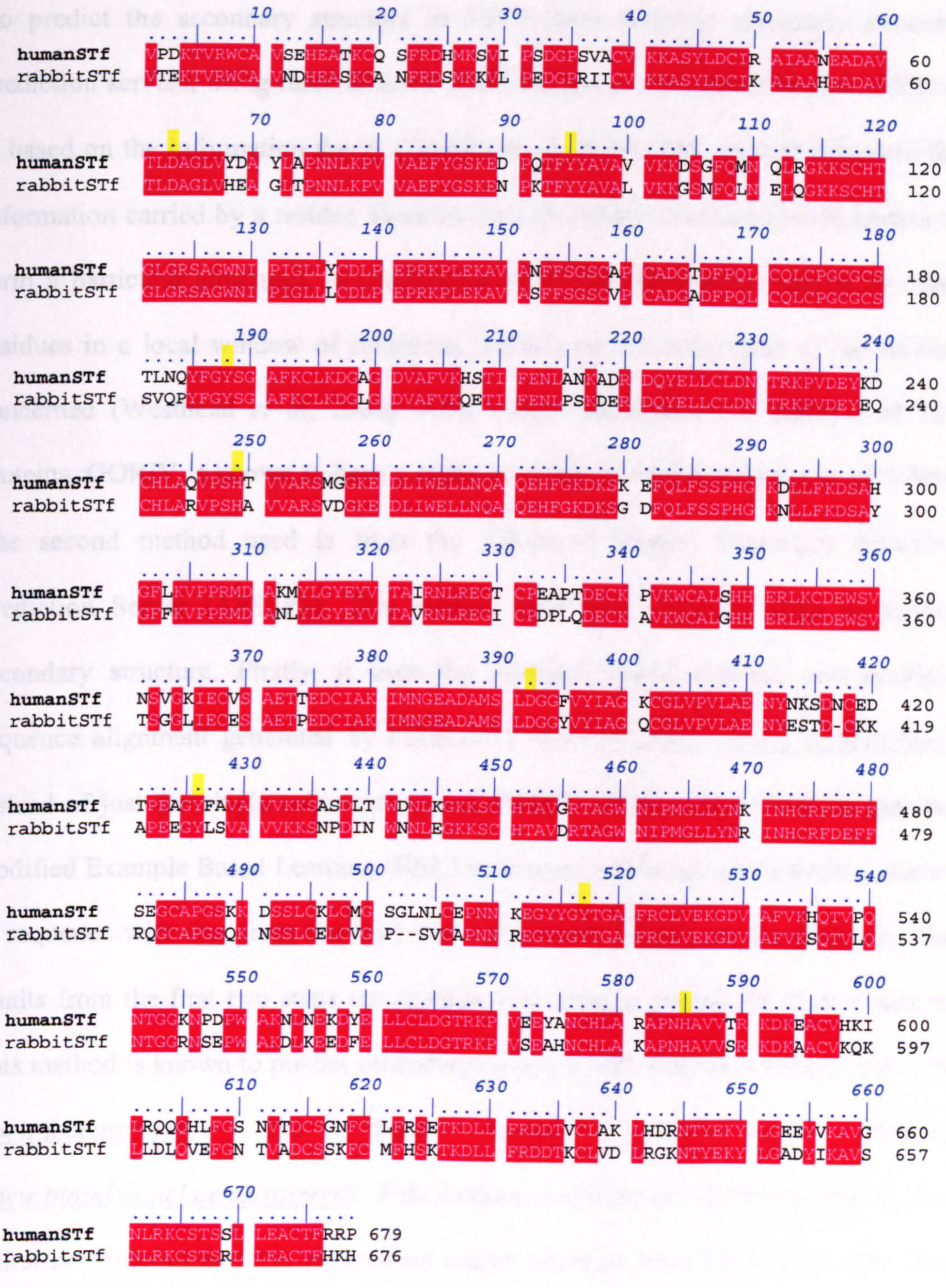


Figure 3.4. Sequence alignment of hSTf and rSTf generated by ClustalW. Sequence in red highlights regions of 100% identity. The eight iron-binding residues for both proteins are marked above in yellow.

3.1.2. Secondary Structure Prediction

3.1.2.1. Secondary Structure Prediction Overview

To predict the secondary structure of hSTf, three different secondary structure prediction servers, using three different methods, were used. The first one, GORIV, is based on the information theory (Garnier *et al*, 1978). This method considers the information carried by a residue about its own secondary structure (its tendencies to form a particular structure), in combination with the information carried by other residues in a local window of seventeen residues on the either side of the residue concerned (Westhead *et al*, 2002). After cross validation on a database of 267 proteins, GORIV is known to have a mean accuracy of 64.4% (Garnier *et al*, 1996). The second method used is from the Advanced Protein Secondary Structure Prediction Server (APSSP2). This method uses three different steps to predict secondary structure. Firstly, it uses the standard neural network and multiple sequence alignment generated by PSIBLAST *(biology.wustl.edu/gcg/psiblast.html)* instead of just the single sequence as with GORIV. In the second step, it uses the modified Example Based Learning (EBL) technique, which compares short segments of sequence with that from known protein structures. Finally, in the third step, the results from the first two steps are combined in order to predict the final structure. This method is known to predict secondary structure with a mean accuracy of 82.9% *(www.forcasp.org)*. The third method used was PSIpred *(www.bioinf.cs.ucl.ac.uk/psipred)*. This method is similar to APSSP2 in that it uses neural networks to perform analysis on output obtained from PSIBLAST. The data collected from four separate neural networks is averaged out in the prediction process to increase accuracy (~78% accurate) (McGuffin *et al*, 2000).

3.1.2.2. Secondary Structure Prediction Analysis

Figure 3.5a, shows the sequence of hSTf aligned with rSTf and the results from the three secondary structure prediction programs discussed above. The secondary structure of rSTf (IJNF) is also shown in this figure. As both hSTf and rSTf are highly homologous, it is reasonable to suggest that these two molecule's secondary structure will be similar. When comparing the predictions from the three methods, they all share a high degree of similarity, especially with APSSP2 and PSIPred. In the majority of cases, these predictions are also homologous with the actual secondary structure of rSTf. This is consistent with the idea both molecules are very closely related in evolutionary terms and therefore should have a similar secondary structure.

However, this does not always have to be the case as two identical sequences can sometimes have different secondary structures due to long-range interactions (Kihara, 2005). This means that such methods as APSSP2 and PSIPred, despite their accuracy in predictions, could actually skew the results. The reason for this is because these two methods use PSIBLAST to predict the secondary structure by analysing the secondary structure of molecules similar to the query sequence. As PSIBLAST is similar to the BLAST search engine used in section 3.1.1.1 there is high probability that rSTf would therefore have been used in the prediction processes. This is the reason why GORIV was also used as this method produces a prediction purely from the amino acid sequence. Even so, as discussed before, APPSP2 and PSIPred have both been shown to be more accurate than GORIV.

a)

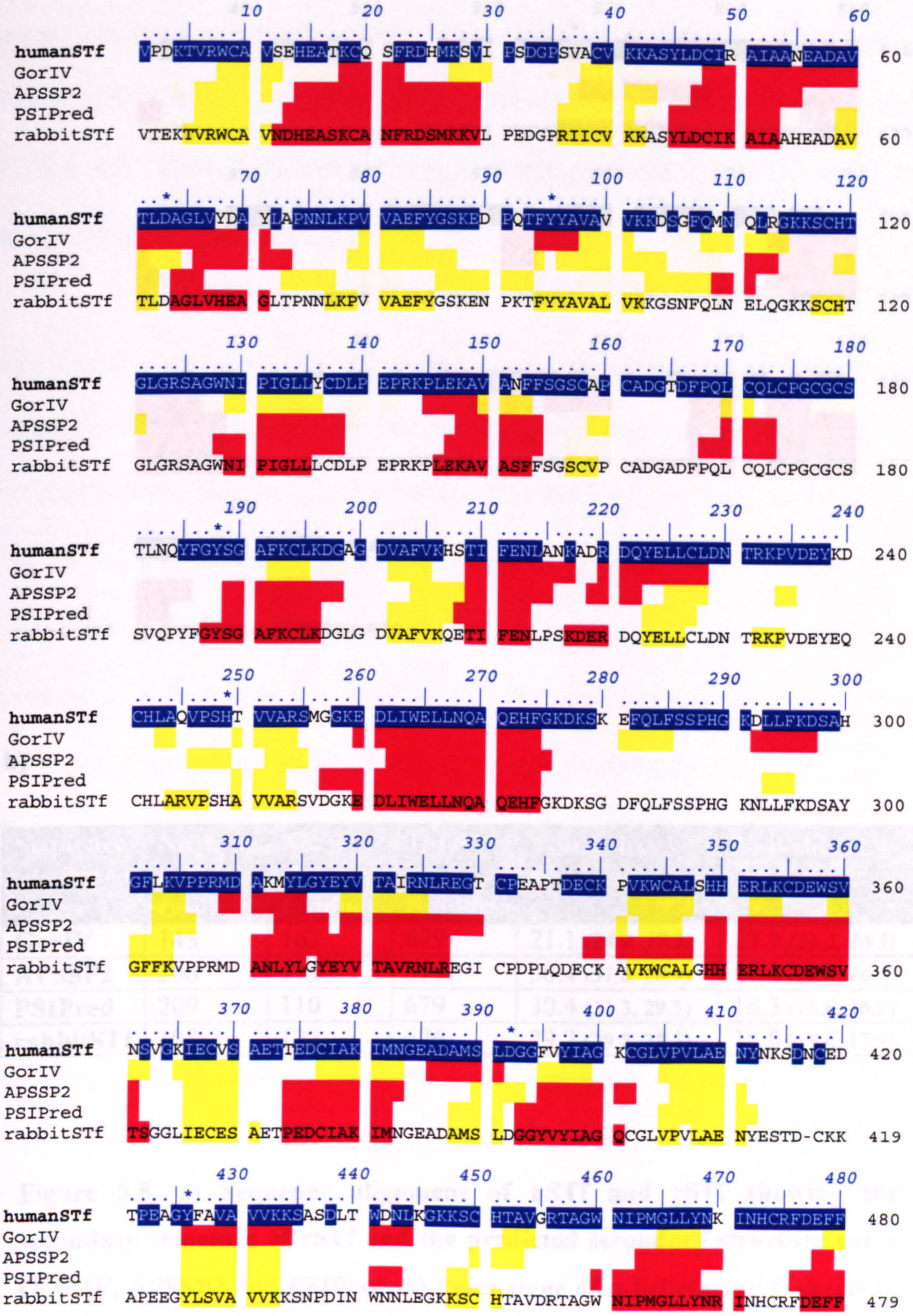
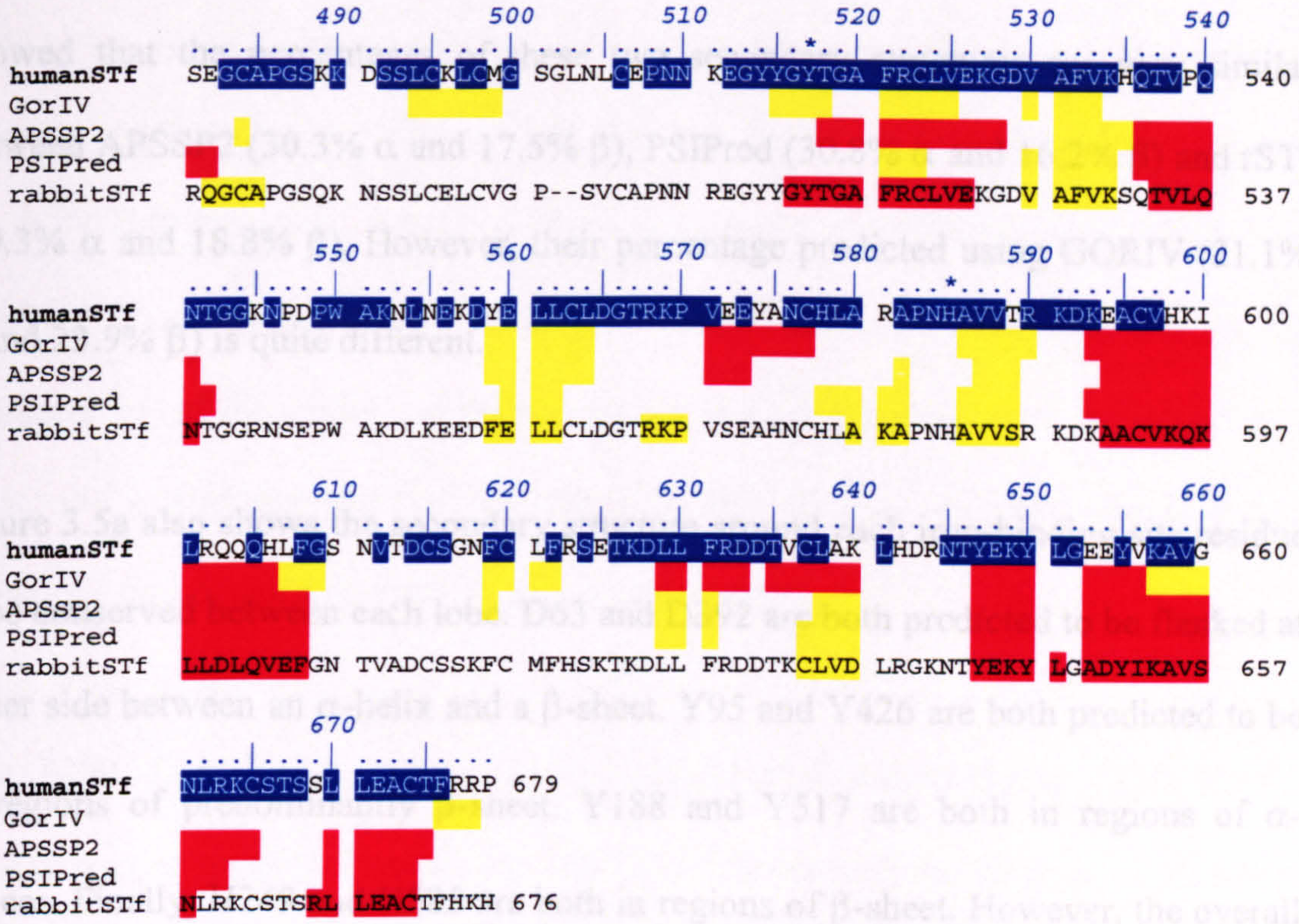


Figure continues on the next page....

Figure Continued....



b)

Sequence	No. α-helix residues	No. β-sheet residues	Total No. residues	% α-helix *(N-lobe, C-lobe)*	% β-sheet *(N-lobe, C-lobe)*
GorIV	143	162	679	21.1 *(24.9, 17.3)*	23.9 *(21.3, 26.3)*
APSSP2	206	119	679	30.4 *(31.8, 28.9)*	17.5 *(19.0, 16.1)*
PSIPred	209	110	679	30.4 (31.3, 29.5)	16.3 (16.6, 16.0)
rabbitSTf	198	127	676	29.2 (29.8, 28.5)	18.8 (19.6, 17.9)

Figure 3.5. a) Sequence alignment of hSTf and rSTf showing the secondary structure of rSTf and the predicted secondary structure using GORIV, APSSP2 and PSIPred. b) Percentage of α-helices and β-sheets in each prediction and for rSTf. Sequence highlighted in blue is sequence that is homologous with rSTf. Red regions: helix (α- or 3_{10}) and yellow regions: β-sheet. * indicates iron-binding residues

To be able to compare the three predictions with the rSTf structure, the percentage of α-helices and β-sheets were calculated for each one (Figure 3.5b). These results showed that the percentages of these two secondary structures are very similar between APSSP2 (30.3% α and 17.5% β), PSIPred (30.8% α and 16.2% β) and rSTf (29.3% α and 18.8% β). However, their percentage predicted using GORIV (21.1% α and 23.9% β) is quite different.

Figure 3.5a also shows the secondary structure around each iron-binding site residue to be conserved between each lobe. D63 and D392 are both predicted to be flanked at either side between an α-helix and a β-sheet. Y95 and Y426 are both predicted to be in regions of predominantly β-sheet. Y188 and Y517 are both in regions of α-helices. Finally, H249 and H585 are both in regions of β-sheet. However, the overall percentage of α-helix and β-sheet does vary between lobes to as much as 8%, showing that the secondary structure of both lobes is significantly different.

3.1.3. Homology Modelling

From the results obtained in section 3.1.1. and section 3.1.2, rSTf was found to be the best candidate for a template to create a homology model of hSTf due to its high degree of homology with hSTf, at both the primary and predicted secondary structure levels. To create this model, a Silicon Graphics Indy Workstation and Modeller 6 (salilab.org) were used. The core modelling procedure begins with the alignment of the sequence to be modelled (target) with related known 3D structures (template). An alignment of these two sequences is given as an input to the program, with the output being a 3D model.

3.1.3.1. Creation of hSTf Model

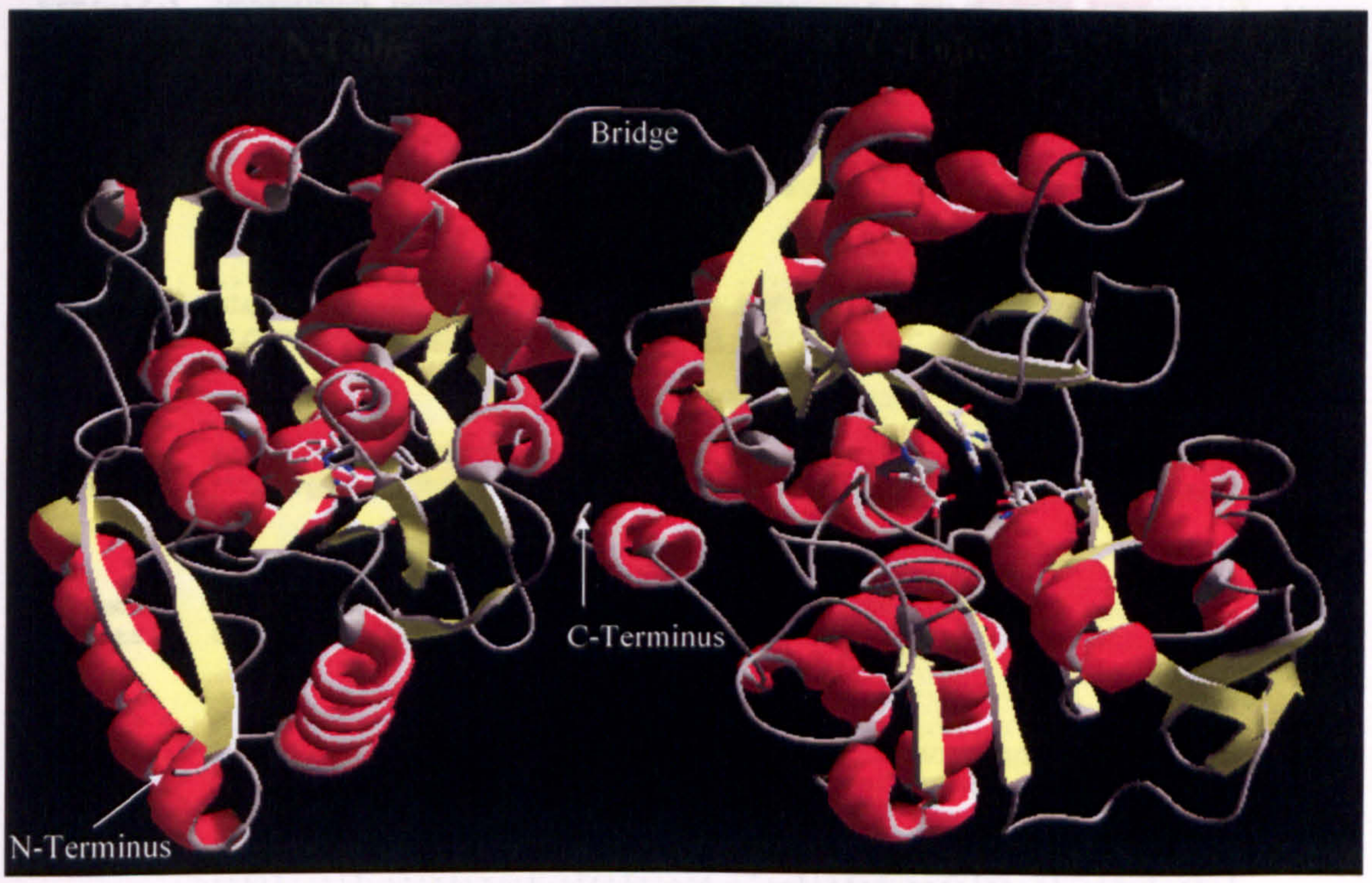
A model of hSTf was created as described in the Modeller 6 manual (Sali *et al*, 2001). The three input files to this program included the alignment (PIR database format) of hSTf with rSTf (Figure 3.4), the PDB file of rSTf (IJNF) and the sequence of hSTf (NP_001054). After the fully automated modelling program had been processed, two output files had been created. The first of these is a PDB file (Appendix I) of the model and the second one is a log file describing the information about the run (Appendix I). This latter file was checked to ensure no additional violations had occurred during the modelling process.

The PDB file created was viewed using Deep Viewer/Swiss PDP Viewer (us.expasy.org/spdbv/) (Figure 3.6). This figure shows a ribbon diagram of the model with the β-sheets highlighted in yellow and the α-helices highlighted in red. A diagram of the two iron-binding sites is also shown with the eight iron-binding residues labelled. The percentage of α-helices and β-sheets of the model, shown in Figure 3.7, are quite similar to the percentages obtained for rSTf and hSTf predictions in section 3.1.2.

3.1.3.2. Model Validation

The first validation test was to create a Ramachandran plot of the hSTf model and of rSTf using Deep Viewer/Swiss PDB Viewer (Figure 3.8). Three main regions can be seen on these plots. The top left hand corner of these plots represents β-sheet, below this region represents right handed α-helix and in the top right quarter represents left handed α-helix. Each white cross or square represents a particular residue within the 3D model and its position according to the phi and psi angles that it possesses. Areas

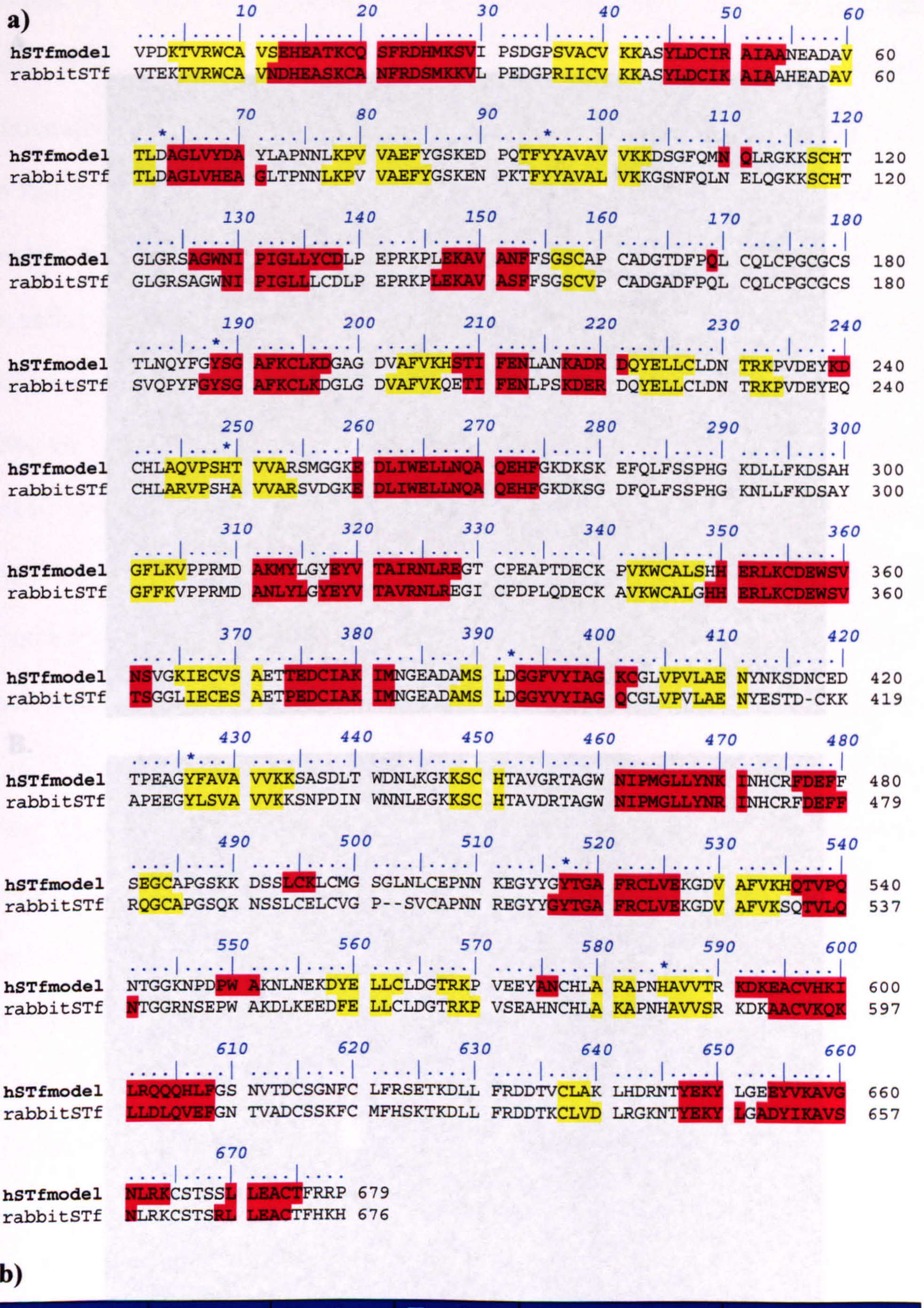
A.



B.

Figure 3.6. The 3D model of hSTf created using Modeller 6 and a Silicon Graphics Indy Workstation and viewed using Deep Viewer / Swiss PDB Viewer. Red: α-helix. Yellow: β-sheet. Grey: Random coil. A: A ribbon diagram of the model with the two lobes highlighted. B: A picture of the iron-binding site of each lobe with the eight iron-binding residues labelled.

a)

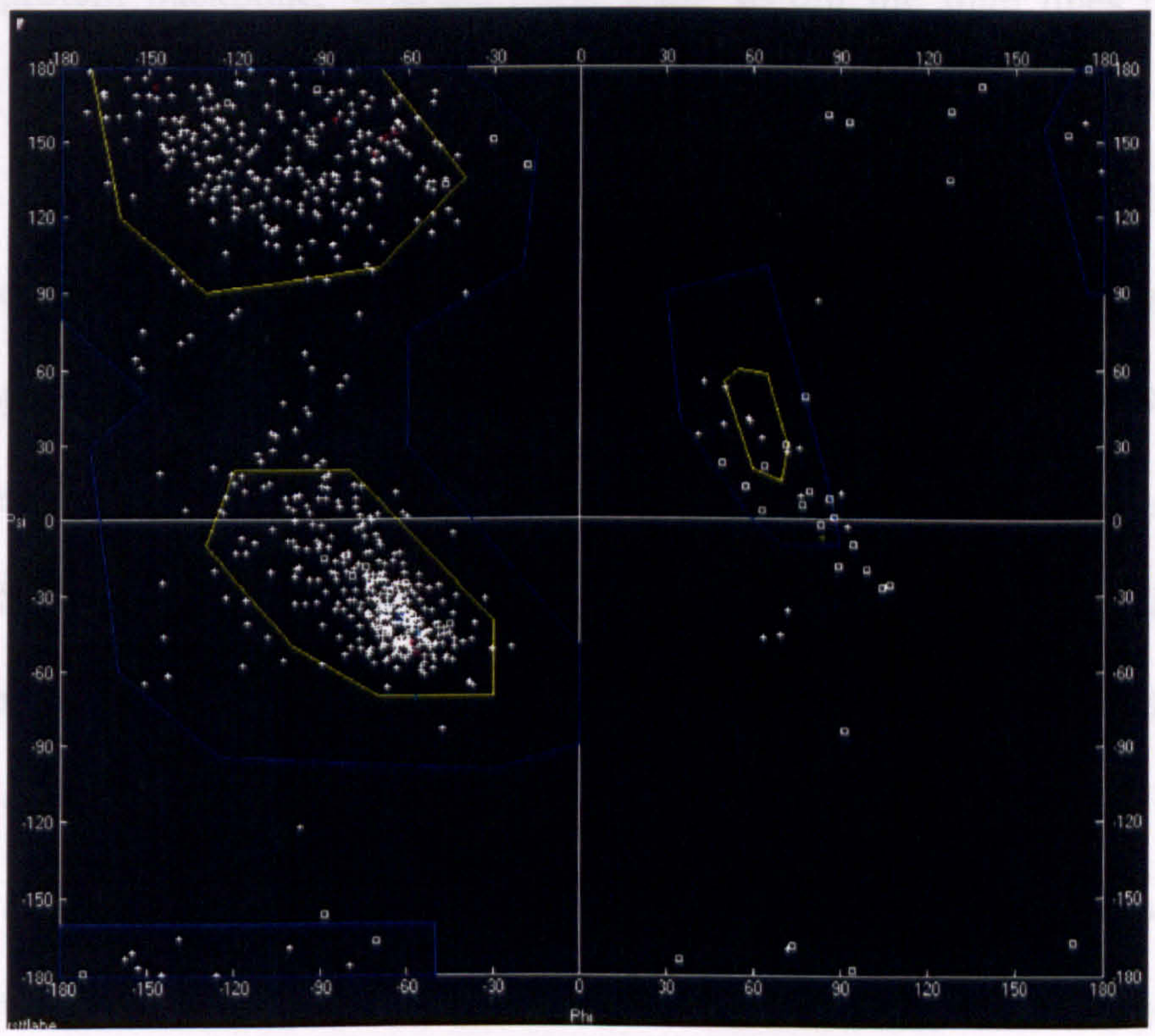


b)

Sequence	No. α-helix residues	No. β-sheet residues	Total No. residues	% α-helix *(N-lobe, C-lobe)*	% β-sheet *(N-lobe, C-lobe)*
hSTfmodel	219	140	679	32.3 *(33, 32)*	20.6 *(22, 19)*

Figure 3.7. a) Sequence alignment of the hSTf model with rSTf, b) Percentage of α-helices and β-sheets. Sequence highlighted in red: α-helices and highlighted in yellow: β-sheets. * indicates iron-binding residues.

A.



B.

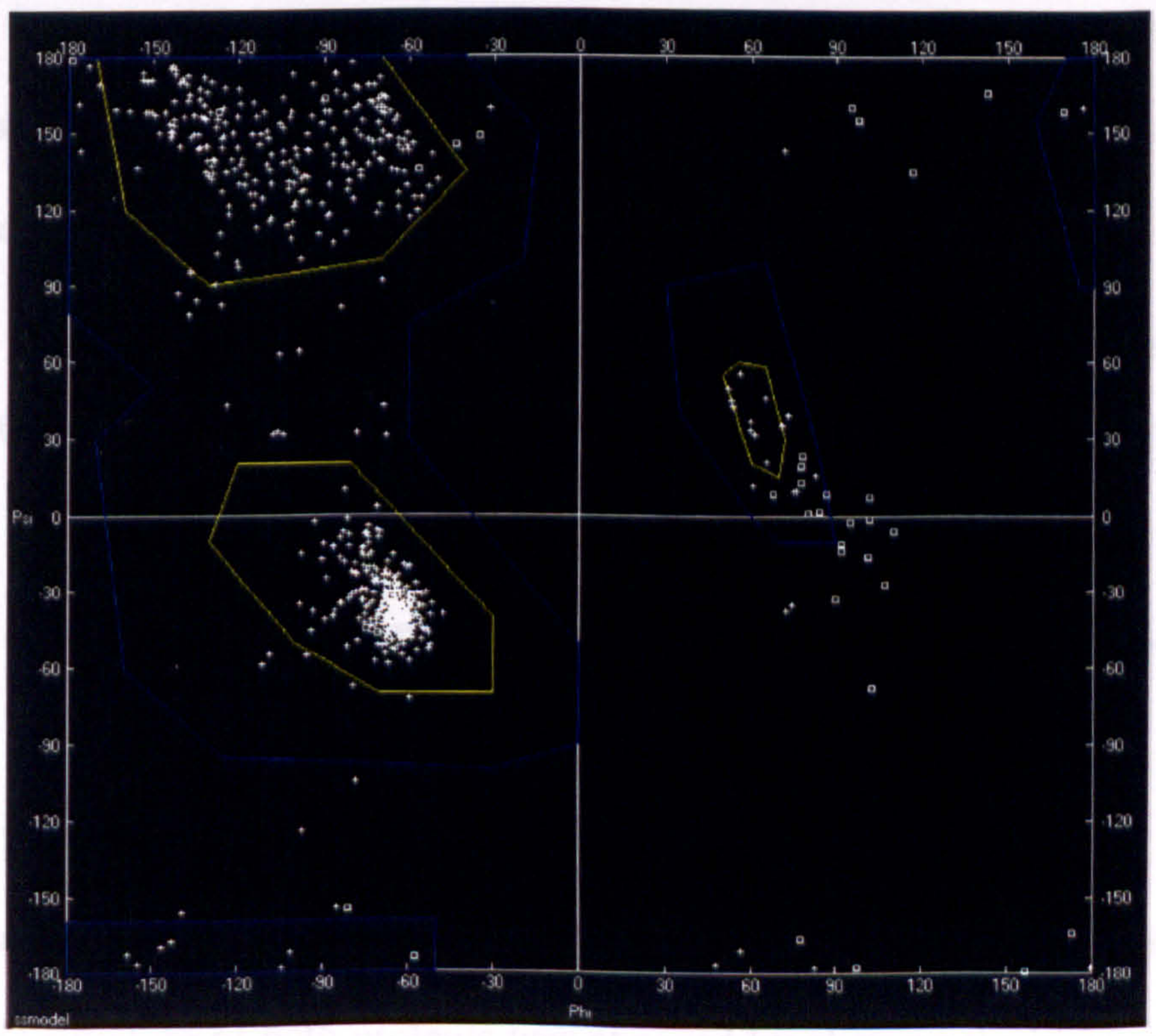


Figure 3.8. Ramachandran Plots created using Deep Viewer / Swiss PDP Viewer. A: Plot of rSTf PDP (IJNF). B: Plot of the hSTf model. Square: Glycines. Crosses: all other residues.

enclosed by a yellow line represent the most favoured phi and psi angle combinations within the protein backbone, whereas those areas within the blue lines represent additionally allowed *phi* and *psi* angle combinations. Anything outside either of these two regions represents residues with disallowed *phi* and *psi* angle combinations. To calculate the frequency of residues in each region, an on-line program known as Procheck (www.biochem.ucl.ac.uk/~roman/procheck) was used. A score of 91.4% of hSTf residues were found to be located in the most favoured regions compared with 85.6% for rSTf. Of the remaining portion, 8.2% of hSTf reside in the additionally allowed regions compared with 13.9% for rSTf. The remaining residues for both proteins reside in the disallowed regions. What is noticeable is that hSTf has more of its residues in the most favoured regions, even though the data for rSTf is from a crystal structure. There are two possible reasons for this. Firstly, the data collected for rSTf is from a protein crystal, and therefore describing a static state of the protein. This means that the protein may not be in its most natural configuration and therefore some bond angles could be strained. Secondly, within the rules used by Modeller to create a model, the ones including the optimum phi and psi angles will tend to lead to a high percentage of residues located into the preferred Ramachandran plot regions. However, in general, the majority of residues for both proteins reside in the favoured regions and when comparing the two proteins, a lot of similarity can be seen.

Knowing that the Ramachandran plots of both proteins are similar, the similarity in the overall 3D structure of both proteins was checked by aligning the hSTf model to the rSTf template using Deep Viewer/Swiss PDP Viewer (Figure 3.9). By viewing

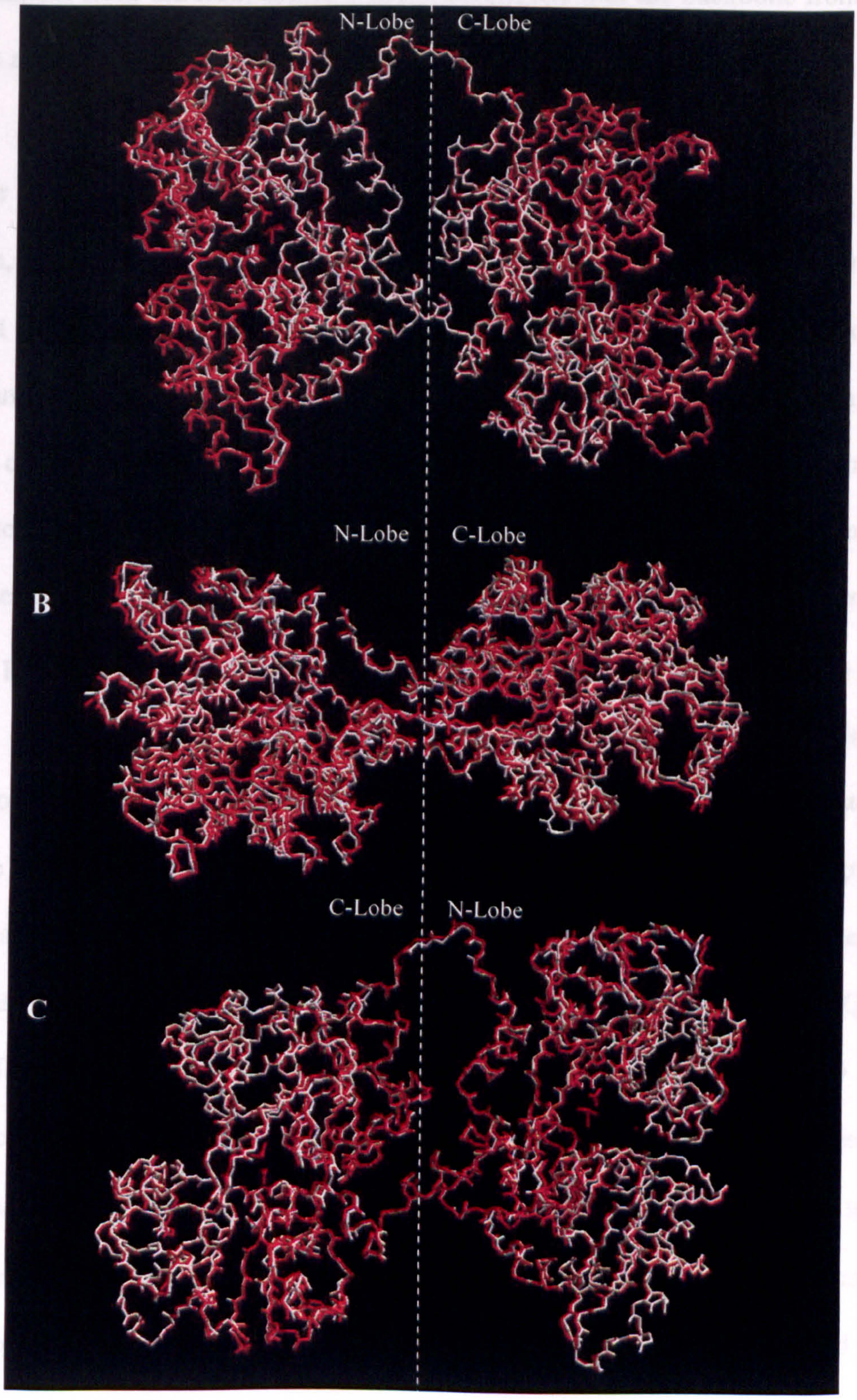


Figure 3.9. Structural alignment of rSTf (IJNF) with the hSTf model. The alignment of the two molecules without their side-chains was viewed using Deep Viewer / Swiss PDB Viewer. Red: rSTf and white: hSTf model. A-C: different views of the alignment.

the alignment from different angles it is possible to see that the backbone from both proteins are quite superimposed.

In order to validate the model further, another on-line program was used. This program, known as WHATIF (biotech.ebi.ac.uk:8400/cgi-bin/sendquery), runs 54 different structural analysis checks on an inputted PDB file. Both the PDB file of the model and also the rSTf (IJNF) were analysed using this program. For the majority of these checks very few errors were noticed. The first error was showing abnormally short interatomic distances, or atomic bumps. Two atoms are said to bump when they are closer than the sum of their Van der Waals radii minus 0.4Å. For hydrogen-bonded pairs, a tolerance of 0.55Å is used. The atomic pairs that bump in both proteins are shown in the Appendix I.4 and I.5. 418 atomic pairs in the model are shown to bump and 180 atomic pairs in rSTf. However, the vast majority of atomic pairs are labelled as BF, meaning that the B-factor average for both atoms is greater than 80 and therefore there is a high probability the atoms are not there. The second error was histidine, asparagine, glutamine side-chain flips. Six residues (Q18, Q109, H205, N323, H533, H640) were shown through hydrogen-bonding analysis to form energetically more favourable hydrogen bonds if the terminal group is rotated by 180°. The remaining errors picked up by WHATIF program were more related to the PDB file than to the model itself. Following these model validation checks, it can be concluded that the accuracy and suitability of the model of hSTf created was acceptable.

3.2. Structure Prediction

Sali *et al*, 1995, stated that with a template structure of more than 40% sequence identity to the target protein, the model is likely to have about 90% of the main-chain atoms modelled with a root mean squared (rms) deviation from the X-ray structure of approximately 1 Å. Therefore, since the sequence identity of hSTf to rSTf (template) is 78% (section 3.1.1.1) the accuracy of the model should be greater than 90%.

3.2.1. Conserved Regions

As discussed in section 3.1.1 a high proportion of residues amongst the Transferrins are conserved. Figure 3.10 highlights the residues of the hSTf model that are found to be conserved amongst the six Transferrins listed in Figure 3.1 and hSTf. From this, important regions of the protein can be determined. From this figure it can be concluded that the conserved residues make up a core skeleton, defining clearly the two domains in each lobe, and the residues that are not conserved, and therefore can be described as variable, surround the outside of the molecule. This is consistent with the transferrins having a similar iron-binding mechanism, but having specificity for a variety of receptors.

3.2.2. Proposed Regions and Residues of Importance

As described in section 1.3.3 all of the iron-binding residues are conserved amongst the majority of the transferrin family, especially between the serum transferrins and lactoferrins. These residues are highlighted in red in Figure 3.10. When visualising the iron-binding residues in relation to secondary structure, all these residues are located, usually towards the end, of α-helix or β-sheets, and never in regions of random coil, or loop (Figure 3.7, section 3.1.3.1). A possible reason for this is so

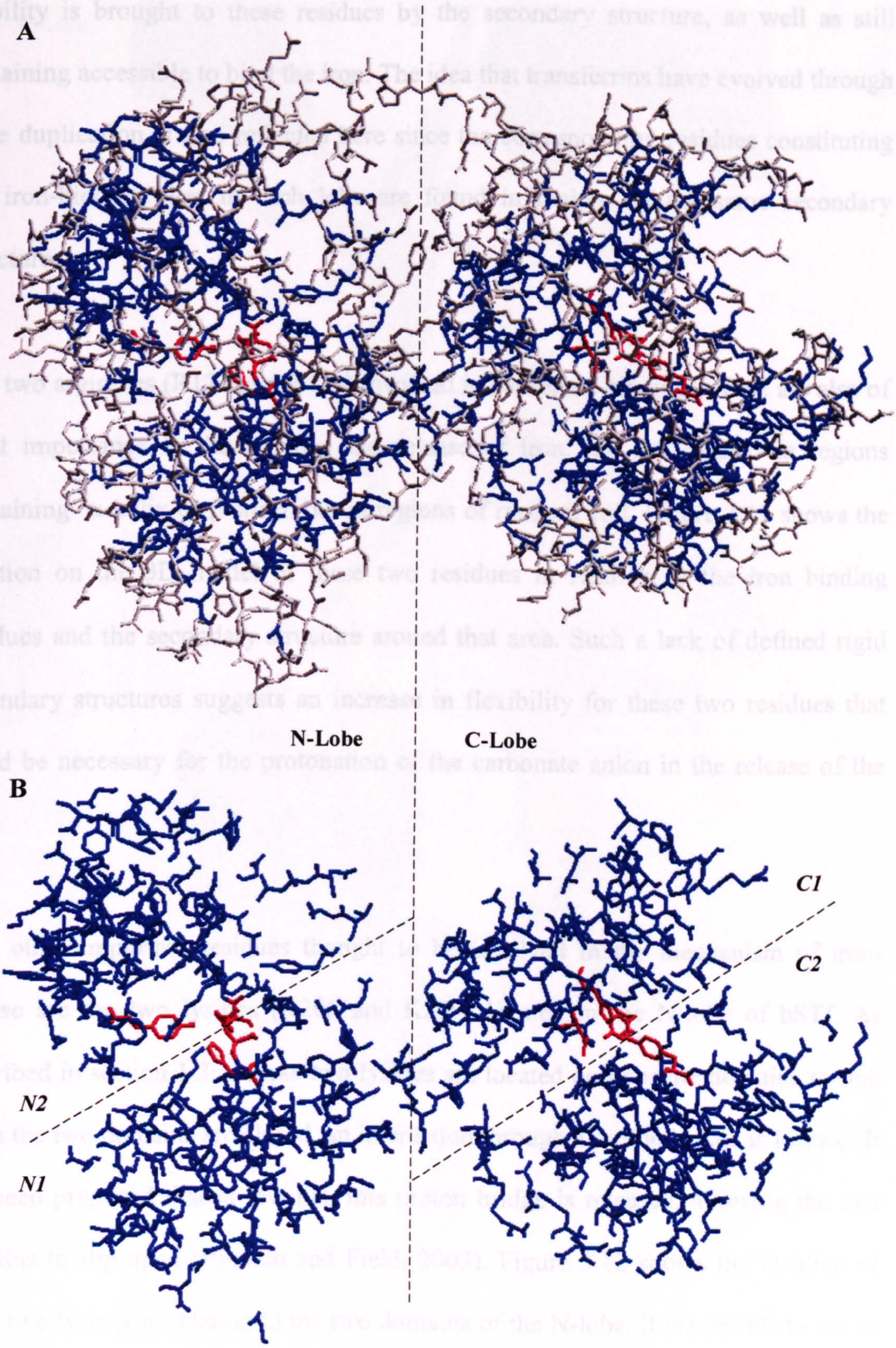


Figure 3.10. PDB image of the hSTf model highlighting residues (blue) that are conserved amongst the six Transferrins listed in section 3.1.1. and hSTf. A: All residues shown. B: Only conserved residues shown. All iron-binding resiudes are conserved and are highlighted in red. Both lobe and the four domains are labelled. (Generated using Deep Viewer / Swiss PDB Viewer)

stability is brought to these residues by the secondary structure, as well as still remaining accessible to bind the iron. The idea that transferrins have evolved through gene duplication is also reflected here since the corresponding residues constituting the iron-binding sites in each lobe are found in regions of the same secondary structure.

The two arginines (R124 and R456) involved in carbonate binding, which are also of great importance in the binding and release of iron, are not located in regions containing α-helix or β-sheet, but in regions of random coil. Figure 3.11 shows the position on the 3D model of these two residues in relation to the iron binding residues and the secondary structure around that area. Such a lack of defined rigid secondary structures suggests an increase in flexibility for these two residues that could be necessary for the protonation of the carbonate anion in the release of the iron.

Two other important residues thought to be involved in the mechanism of iron-release are the two lysines (K206 and K296) located in the N-lobe of hSTf. As described in section 1.3.4, these two lysines are located on opposite domains so that when the two domains are closed, an interaction through a proton bridge is formed. It has been proposed that at low pH, this proton bridge is released, allowing the two domains to flip open (Rinaldo and Field, 2003). Figure 3.12 shows the location of these two lysines in relation to the two domains of the N-lobe. It is possible to see in this figure that these two lysines come into close contact when the two domains N1 and N2 are in the closed configuration shown in the model. The distance between the nitrogen and the hydrogen of both amino acid side-chains is shown in this model to

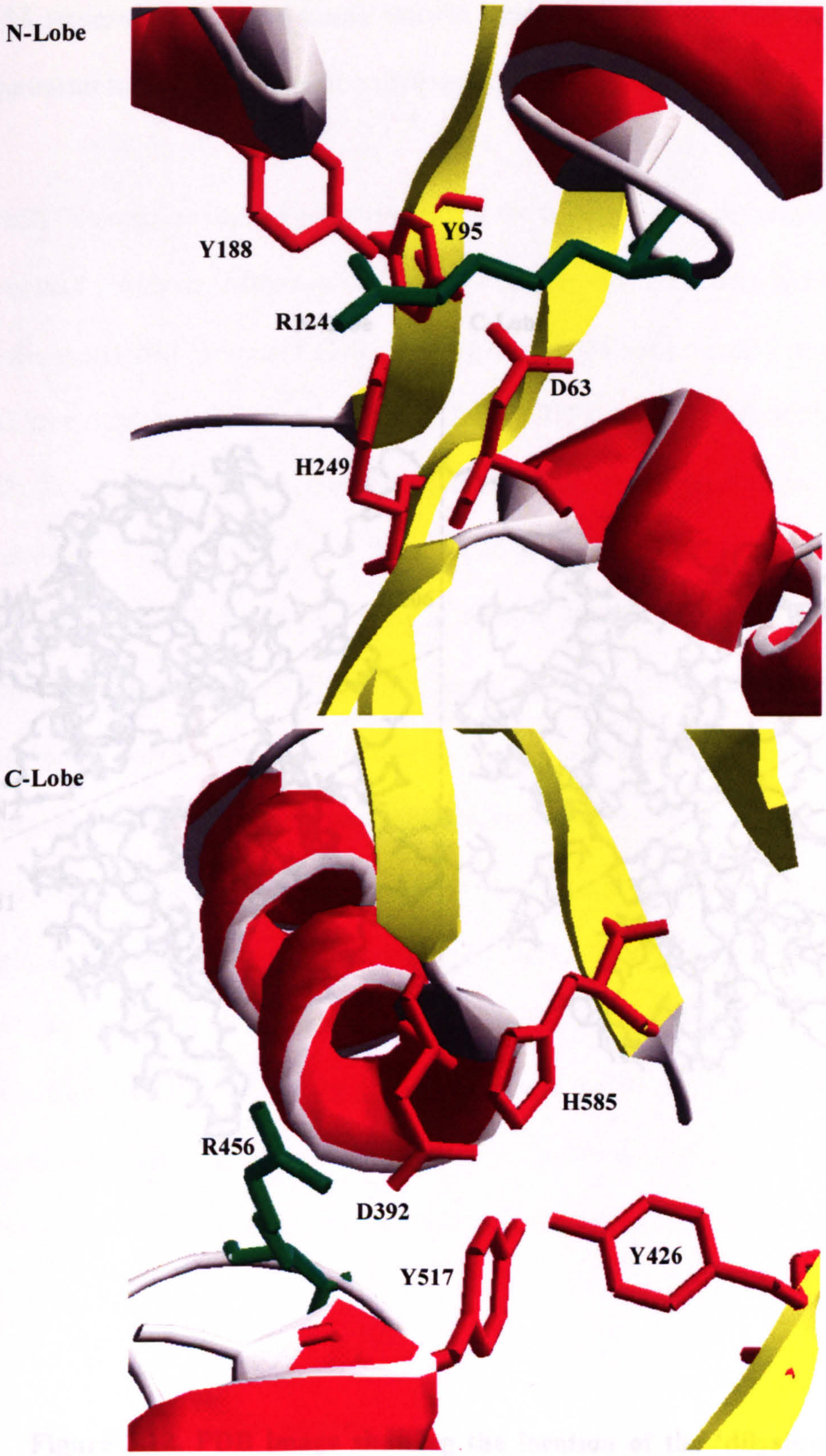


Figure 3.11. PDB image showing the two iron-binding sites of the hSTf model and the secondary structure around those regions. Yellow: β-sheets. Red: α-helix. Grey: random coil, or loop. Red residues: directly involved in iron binding. Green residues: involved in carbonate binding.

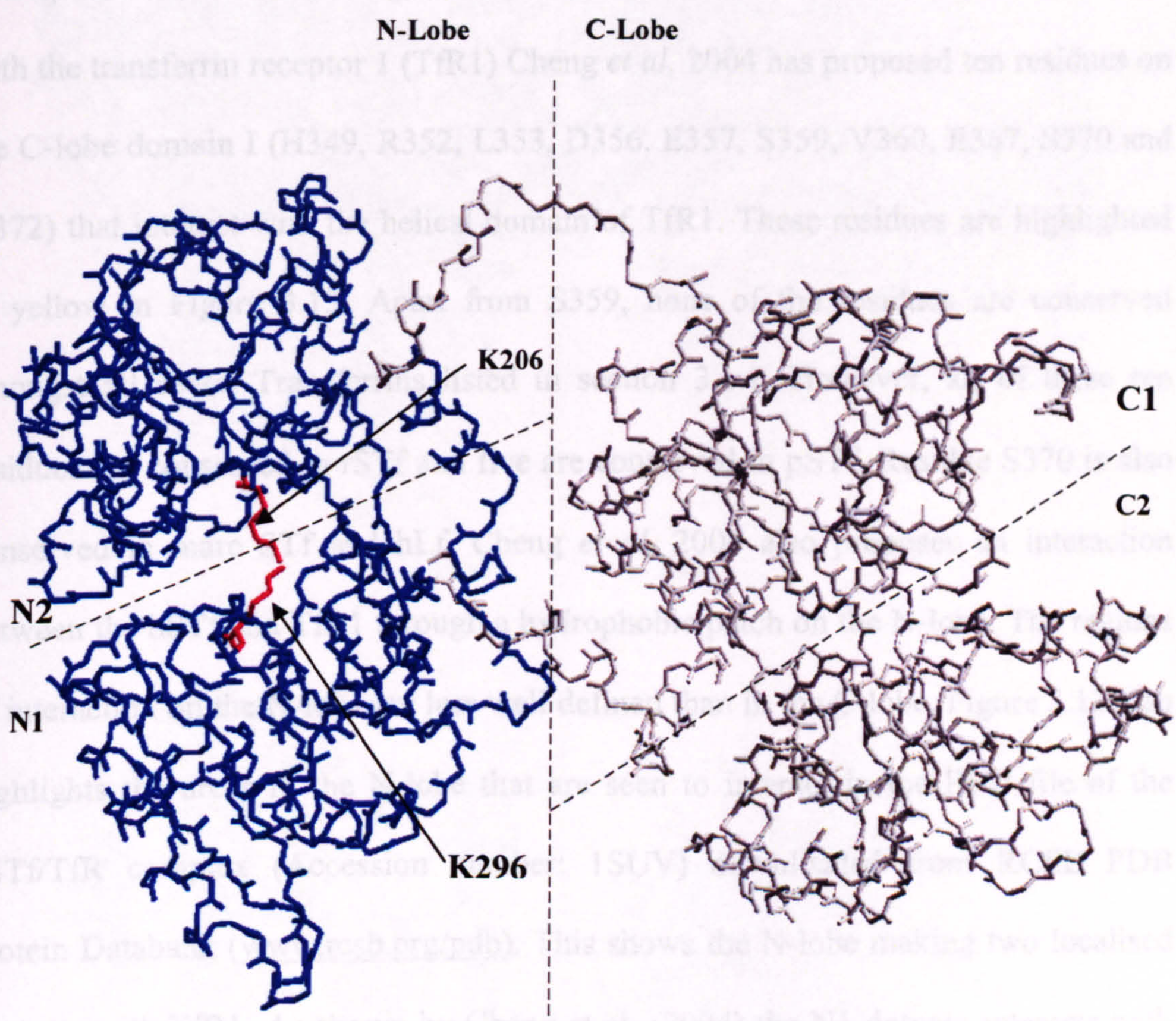


Figure 3.12. PDB image showing the location of the 'dilysine-trigger' in the N-lobe of the hSTf model, minus all side-chains (apart from K206 and K296). The two are highlighted in red, the C-lobe in grey and the N-lobe in blue. The four domains are also labelled (N1, N2, C1 and C2).

be ~3Å (measured using the atomic ruler in Deep Viewer/Swiss PDB viewer) which is equivalent to the typical length of hydrogen bonds (Kyte, 1995).

TfR/hSTf interactions are of great interest for the understanding of receptor-mediated iron uptake into cells. Although it is totally not clear which residues in hSTf interact with the transferrin receptor 1 (TfR1) Cheng *et al,* 2004 has proposed ten residues on the C-lobe domain I (H349, R352, L353, D356, E357, S359, V360, E367, S370 and G372) that interact with the helical domain of TfR1. These residues are highlighted in yellow in Figure 3.13. Apart from S359, none of the residues are conserved amongst all seven Transferrins listed in section 3.1.1. However, all of these ten residues are conserved in rSTf and five are conserved in pSTf. Residue S370 is also conserved in mare STf and hLf. Cheng *et al*, 2004 also proposes an interaction between the hSTf and TfR1 through a hydrophobic patch on the N-lobe. The regions of interaction on the N-lobe are less well defined than in the C-lobe. Figure 3.13 also highlights the areas in the N-lobe that are seen to interact in the PDB file of the hSTf/TfR complex (Accession number: 1SUV) downloaded from RCSB PDB Protein Databank (www.rcsb.org/pdb). This shows the N-lobe making two localised contacts with TfR1. As shown by Cheng et al, (2004) the N1 domain interacts with the helical domain of TfR1 and the N2 domain interacts with the protease-like domain.

3.2.3. Surface Characterisation

The surface-characteristics of any ligand molecule are of great importance, since it is at this interface that interactions with other molecules occur. Not only can this help

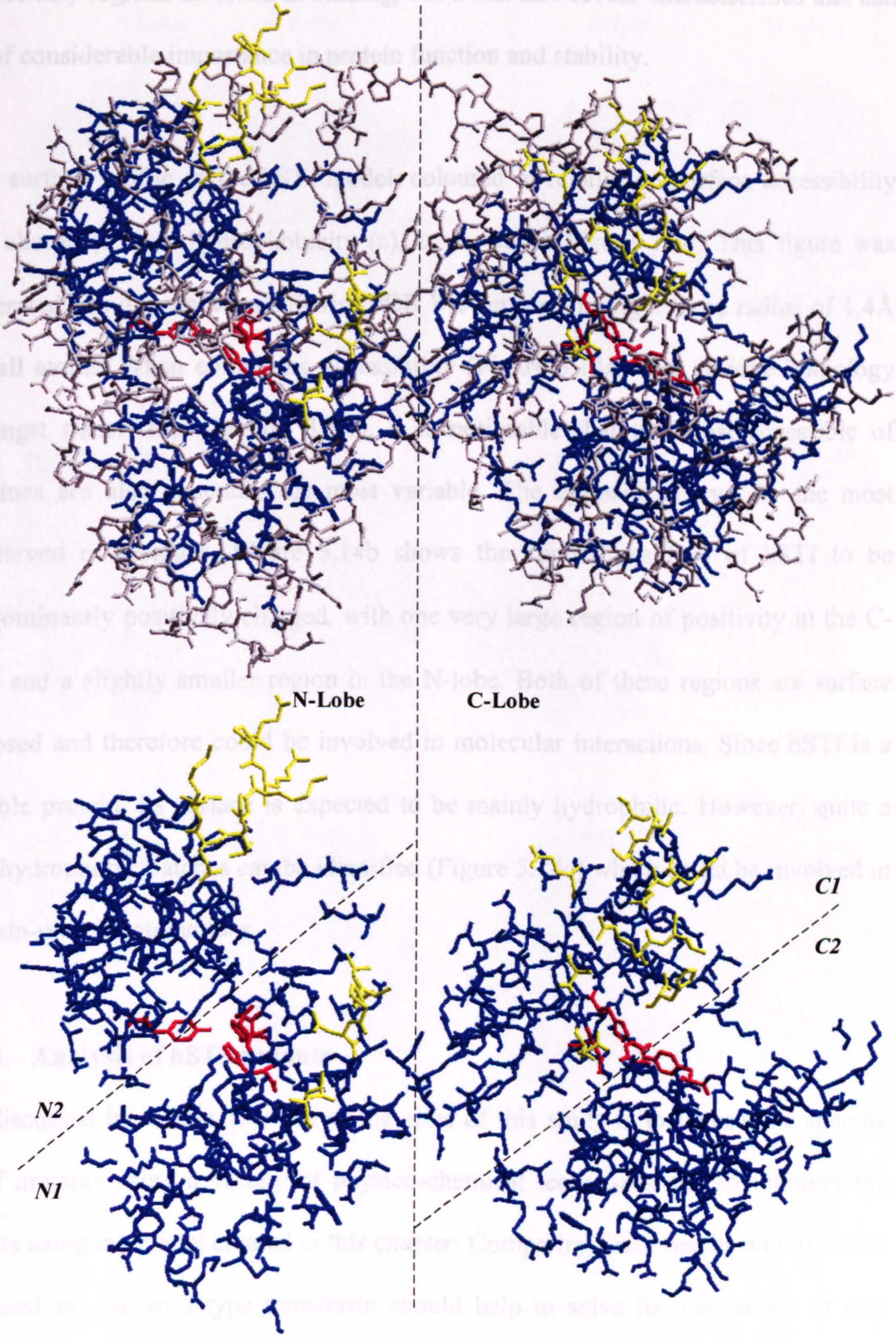


Figure 3.13. PDB image of the hSTf model highlighting regions that are thought to interact with the TfR (yellow) and conserved residues (blue).

A: All residues shown. B: Only conserved residues and TfR interacting residues are shown. Both lobe and the four domains are labelled. (Generated using Deep Viewer / Swiss PDB Viewer)

to identify regions involved in binding, but it can also reveal characteristics that can be of considerable importance in protein function and stability.

The surface profile of the hSTf model, coloured in relation to surface accessibility (a), charge (b) and hydrophobicity (c), is shown in Figure 3.14. This figure was generated using Deep Viewer/Swiss PDB Viewer, presuming a fixed radius of 1.4Å for all atoms. When comparing accessibility (Figure 3.14a) with residue homology amongst transferrins (section 3.1.1), it is noticeable that the most accessible of residues are also generally the most variable. The opposite is true for the most conserved of residues. Figure 3.14b shows the surface residues of hSTf to be predominantly positively charged, with one very large region of positivity in the C-lobe and a slightly smaller region in the N-lobe. Both of these regions are surface exposed and therefore could be involved in molecular interactions. Since hSTf is a soluble protein, its surface is expected to be mainly hydrophilic. However, quite a few hydrophobic patches can be identified (Figure 3.14c) which could be involved in protein-protein interactions.

3.2.4. Analysis of hSTf mutants

As discussed in section 1.4, one of the aims of this study is to investigate specific hSTf mutants using a variety of physico-chemical techniques and to interpret the results using the model created in this chapter. Comparing these results with the ones obtained for the wild type transferrin should help to solve its mechanism of iron uptake and release by transferrin.

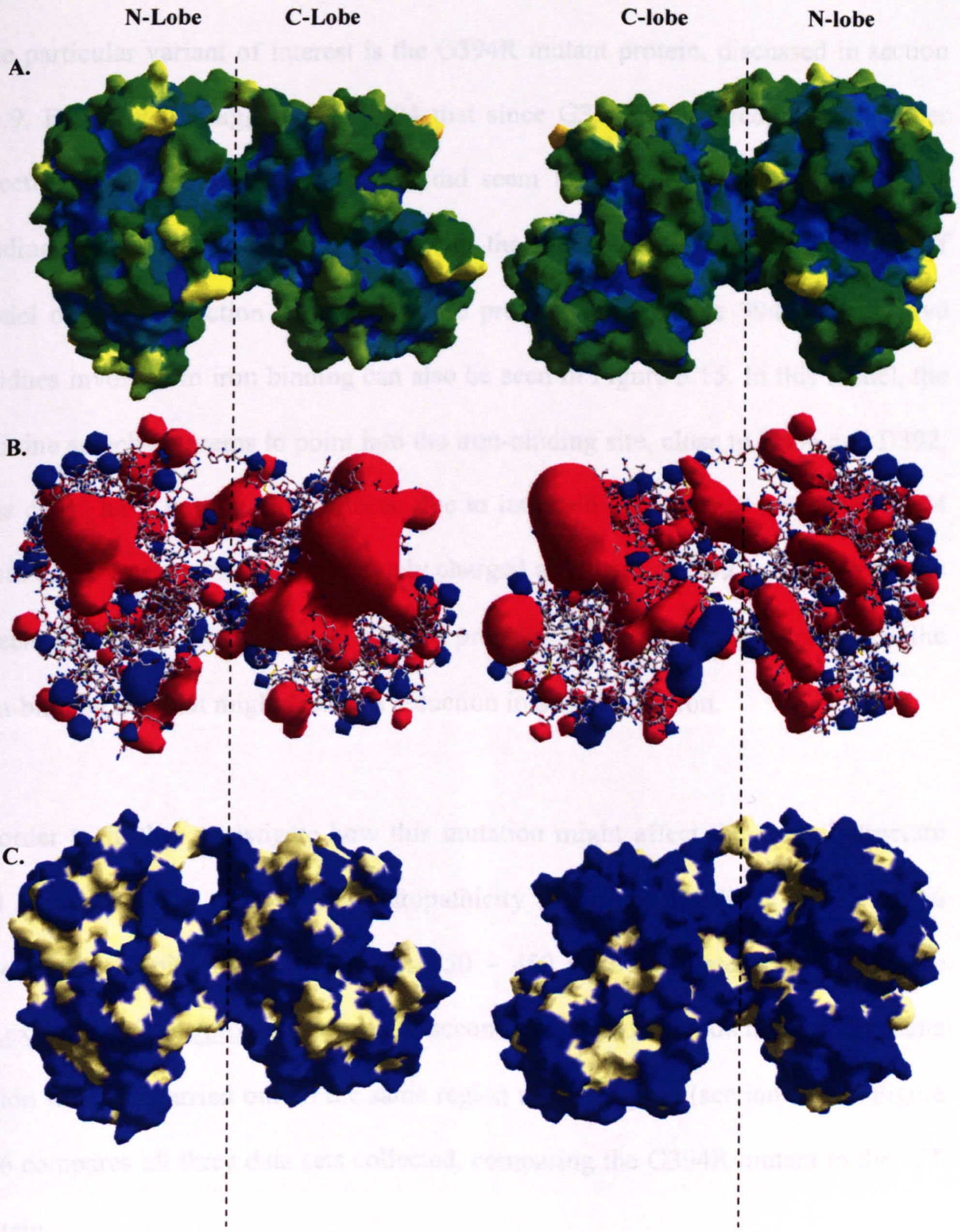


Figure 3.14. PDB image showing the surface accessibility, charge and hydrophobicity of the hSTf model. A: Surface accessibility (yellow most accessible > green > blue less accessible). B: Charge distribution through the protein (red positive, blue negative). C: Hydrophobicity (Blue: Hydrophilic, yellow: hydrophobic). Surface generated using Deep Viewer / Swiss PDB Viewer.

3.2.4.1 G394R Variant

One particular variant of interest is the G394R mutant protein, discussed in section 1.3.9. Evans *et al*, suggested in 1994 that since G394 was buried, it was neither directly involved in iron-binding, nor did seem likely to be involved in receptor binding. These results are consistent with the location of G394 within the hSTf model created in section 3.1.3. The close proximity of residue 394 with the five residues involved in iron binding can also be seen in Figure 3.15. In this model, the arginine sidechain seems to point into the iron-binding site, close to R456 and D392. This could have two possible effects: due to its positively charged sidechain, R394 could either interact with the negatively charged sidechain of D392, or/and repel the sidechain of R456. Either way, this would probably lead to structural alteration of the iron-binding site that might lead to a reduction in affinity for iron.

In order to further investigate how this mutation might affect the overall structure and function of the protein, the hydropathicity (Kyte and Doolittle, 1982) and % accessibility (Janin, 1979) of residues 350 – 450 were calculated using ProtScale (us.expasy.org/tools/protscale.html). A secondary structure prediction of the same region was also carried out on the same region using PSIpred (section 3.1.2). Figure 3.16 compares all three data sets collected, comparing the G394R mutant to the WT protein.

A noticeable difference between both proteins was identified when comparing both secondary structure prediction profiles. In the WT protein, G394 is shown to be on a four-residue section of random coil, sandwiched between two regions of β-sheet. However, in the G394R protein, the secondary structure around this region is

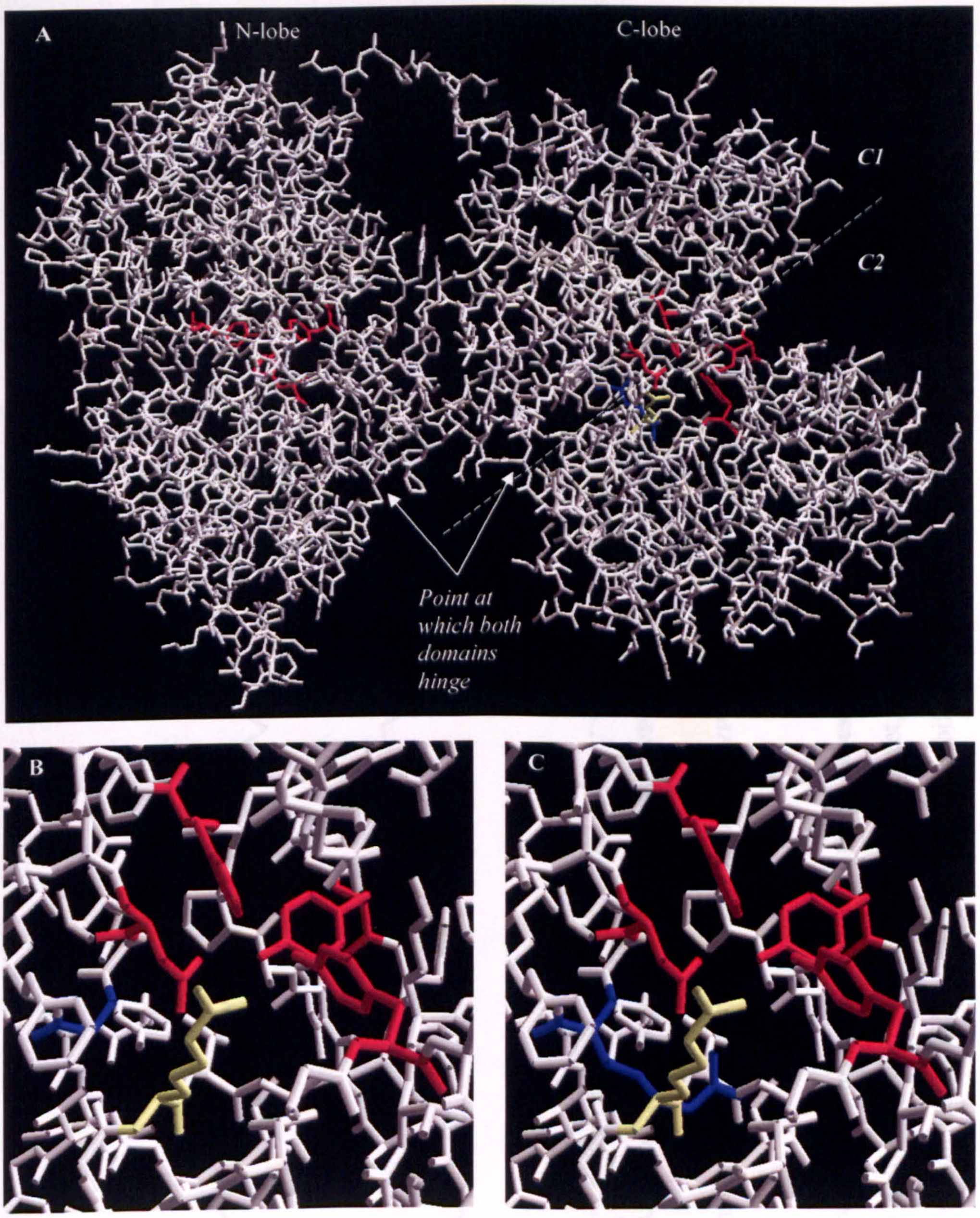


Figure 3.15. Location of G394R mutation within the hSTf model. A: Overview of hSTf and location of the mutant in the C-lobe. B: Location of G394 in relation to the C-lobe iron-binding site. C: Location of R394 in relation to the C-lobe iron-binding site. Red: Residues directly involved in iron-binding. Yellow: Arg456 (involved in binding of the synergistic carbonate anion). Blue: residue 394.

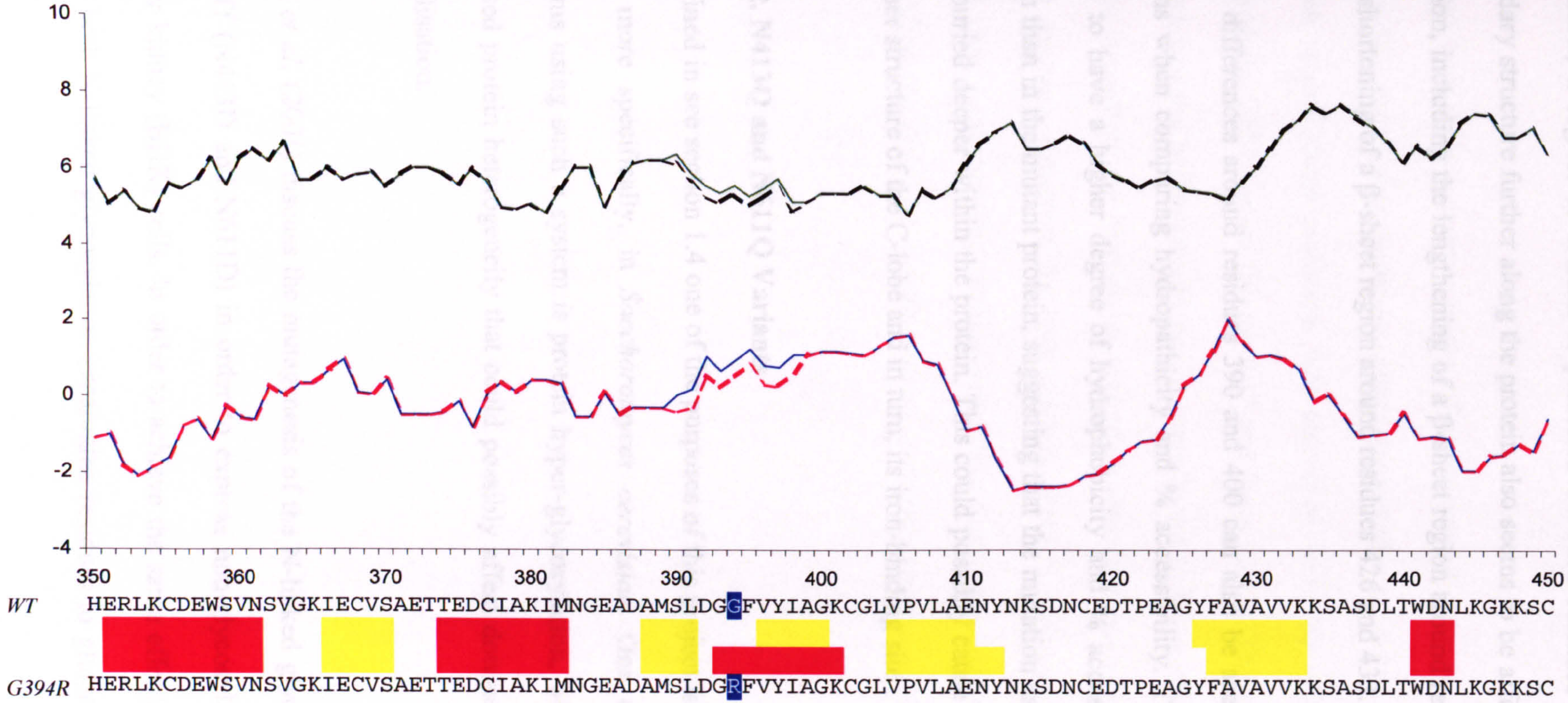


Figure 3.16. Secondary structure prediction, hydropathicity and % accessibility comparison between WT hSTf and G394R hSTf. Red blocks: predicted α-helix, yellow blocks: predicted β-sheet, blue line: WT hydropathicity, red line: G394R hydropathicity, green: WT % accessibility and black line: G394R % accessibility.

dramatically altered. Instead of residue 394 being sandwiched between two regions of β-sheet, a region of α-helix is predicted between residues 393 and 401. The secondary structure further along the protein also seems to be affected by the G394R mutation, including the lengthening of a β-sheet region around residues 405 and 412 and a shortening of a β-sheet region around residues 426 and 433.

Minor differences around residues 390 and 400 can also be seen between the two proteins when comparing hydropathicity and % accessibility. The data shows this region to have a higher degree of hydrophobicity and % accessibility in the WT protein than in the mutant protein, suggesting that the mutation causes these residues to be buried deeper within the protein. This could possibly cause the deformation of the inner structure of the C-lobe and in turn, its iron-binding site.

3.2.4.2. N413Q and N611Q Variants

As defined in see section 1.4 one of the purposes of this project is to express hSTf in yeast, more specifically, in *Saccharomyces cerevisiae*. One of the foreseeable problems using such a system is protein hyper-glycosylation, which might lead to increased protein heterogeneity that could possibly affect downstream analysis and crystallisation.

Mason *et al*, (2001) discuss the mutagenesis of the N-linked glycosylation residues on hSTf (N413D and N611D) in order to express non-glycosylated hSTf in baby hamster kidney (BHK) cells. In order to achieve the same effect, it was decided to mutate both N-linked glycosylation asparagine residues to glutamines (N413Q and

N611Q), which is the residue of choice due to their close similarity, as shown by Dayhoff Table in Appendix I.

To study how these mutations might affect the overall structure and function of the protein, the predicted secondary structure, hydropathicity and % accessibility of the residues around these mutations was compared with the wild type protein, as described for G394 variant (Figure 3.17 and Figure 3.18). The predicted secondary structure and the hydropathicity profile were found to be superimposable for the 3 variants suggesting no changes in secondary structure and hydropathicity for the mutant proteins. However, for both mutants, there is a change in % accessibility around their corresponding regions suggesting that the accessibility of the region surrounding the mutation in both mutants is slightly reduced when compared to the WT.

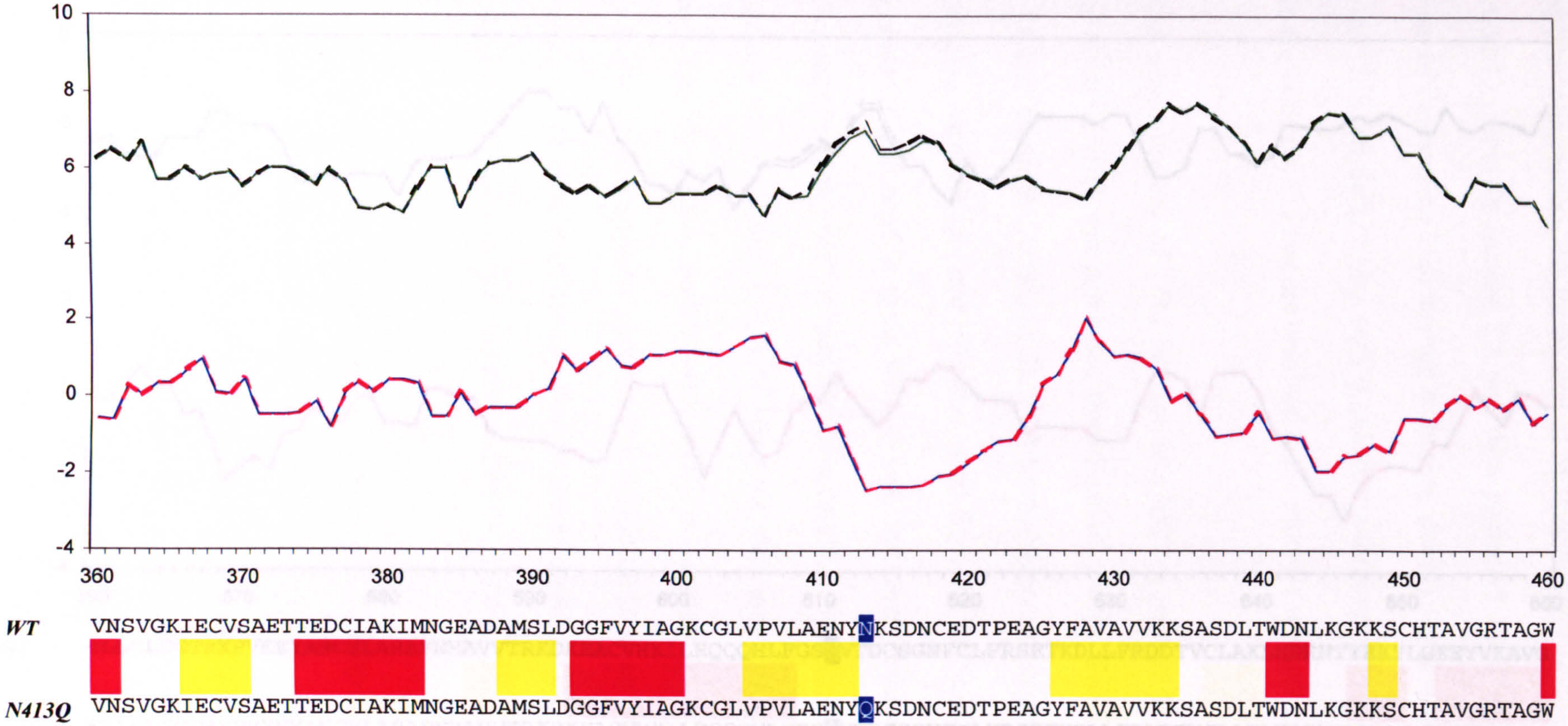


Figure 3.17. Secondary structure prediction, hydropathicity and % accessibility comparison between WT hSTf and N413Q hSTf. Red blocks: predicted α-helix, yellow blocks: predicted β-sheet, blue line: WT hydropathicity, red line: N413Q hydropathicity, green: N413Q % accessibility and black line: WT % accessibility.

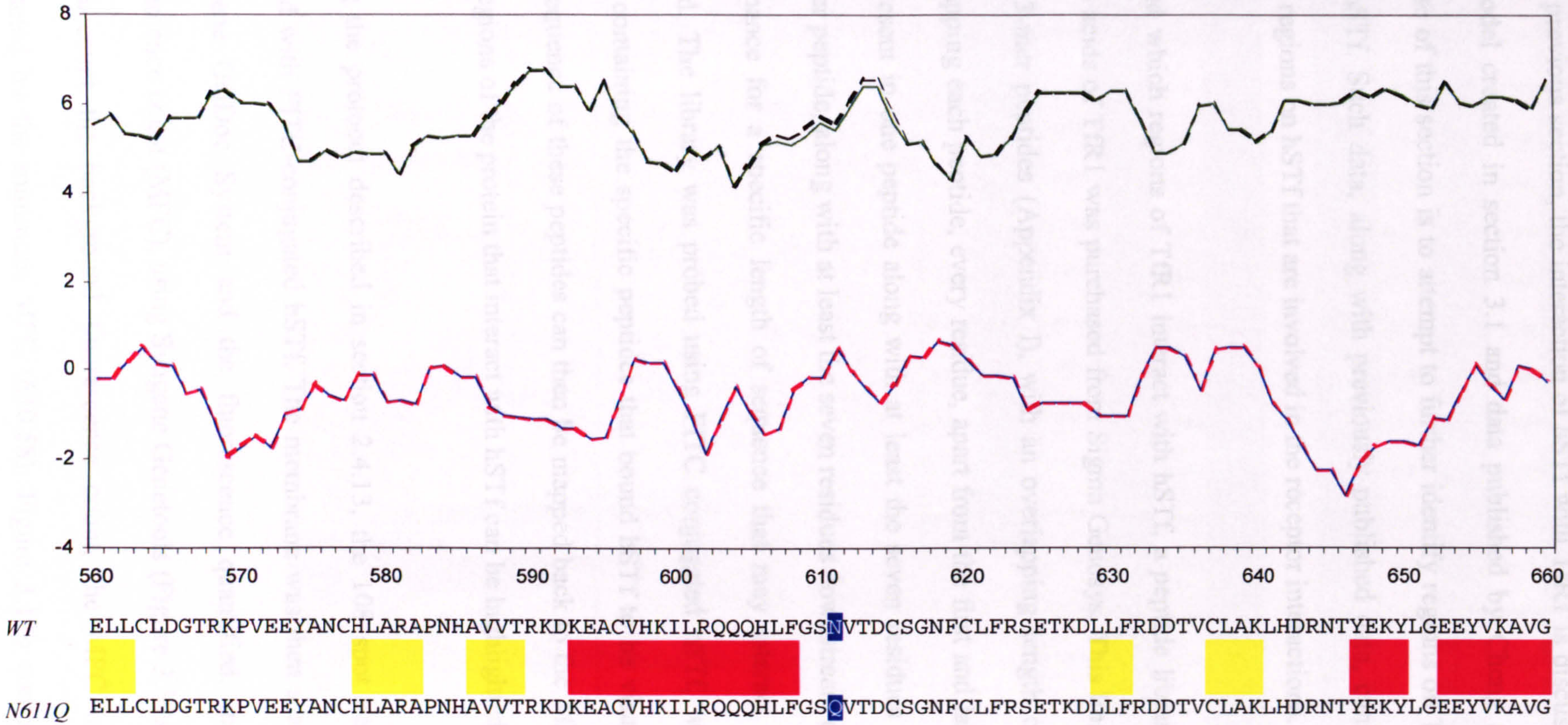


Figure 3.18. Secondary structure prediction, hydropathicity and % accessibility comparison between WT hSTf and N611Q hSTf. Red blocks: predicted α-helix, yellow blocks: predicted β-sheet, blue line: WT hydropathicity, red line: N611Q hydropathicity, green: N611Q % accessibility and black line: WT % accessibility.

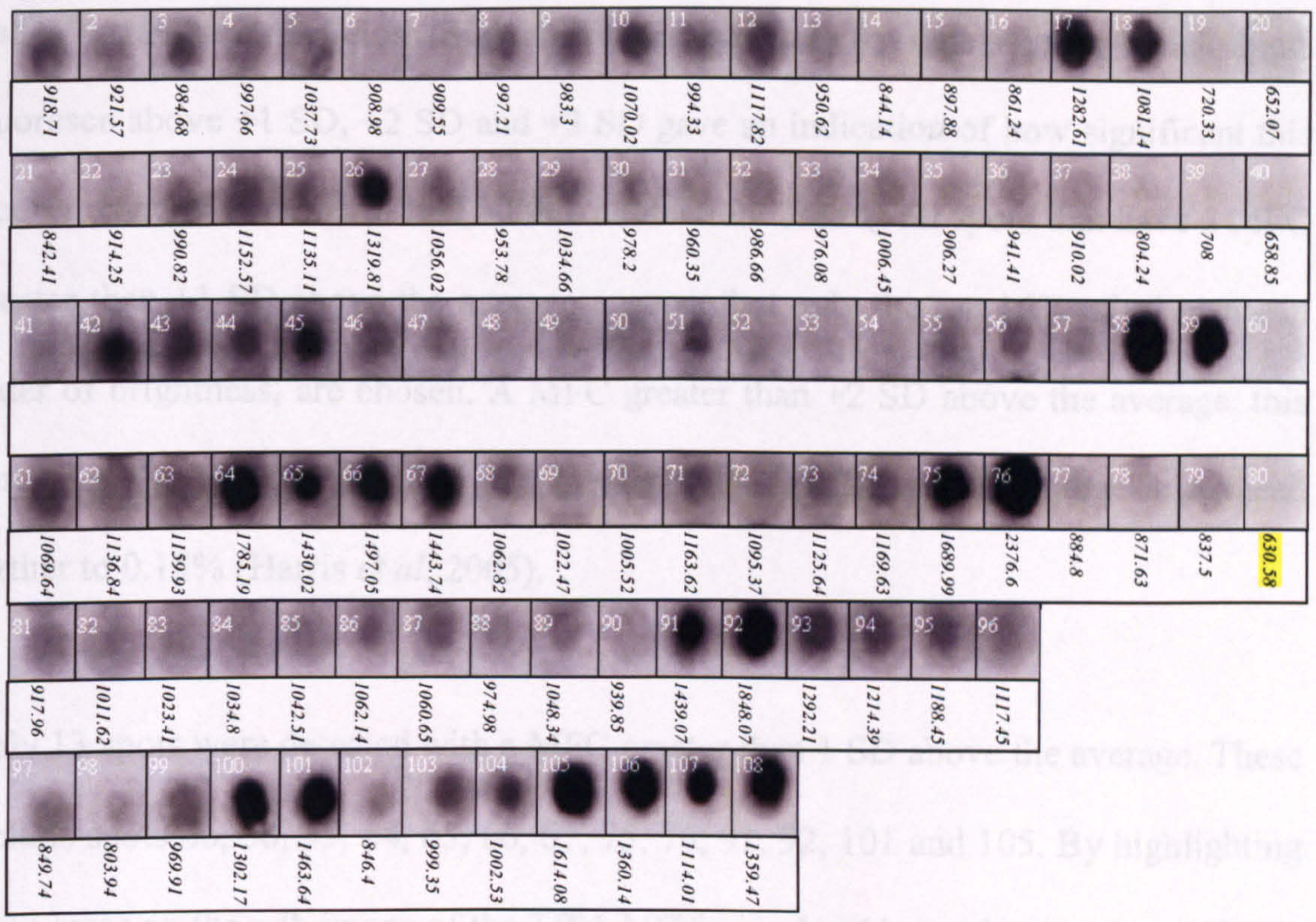
3.3. Transferrin – Transferrin Receptor1 Interactions

In the previous section, the interaction of hSTf with TfR1 is discussed in relation to the model created in section 3.1 and data published by Cheng *et al,* 2004. The purpose of this section is to attempt to further identify regions on TfR1 that interact with hSTf. Such data, along with previously published data, can then be used to locate regions on hSTf that are involved in the receptor interactions.

To find which regions of TfR1 interact with hSTf, a peptide library of the full 760 amino acids of TfR1 was purchased from Sigma Genosys. This library comprises of 108 13-mer peptides (Appendix I), with an overlapping length of 6 residues. By overlapping each peptide, every residue, apart from the first and last seven residues, are present in one peptide along with at least the seven residues upstream, and in another peptide, along with at least the seven residues downstream of it. This reduces the chance for a specific length of sequence that may interact with hSTf being missed. The library was probed using FITC conjugated hSTf, which allowed the spots containing the specific peptides that bound hSTf to be visualised under UV. The sequence of these peptides can then be mapped back to the TfR1 pdb image, so that regions of the protein that interact with hSTf can be highlighted.

Using the protocol described in section 2.4.13, the 108 spot peptide library was probed with FITC-conjugated hSTf. The membrane was then scanned using a UV SynGene GelDoc System and the fluorescence quantified, in terms of mean fluorescence count (MFC), using Syngene Genetools (Figure 3.19a). In order to take into account the background fluorescence count, the MFC of each spot was subtracted by the minimum MFC (630.58). Figure 3.19b compares the adjusted

A.



B.

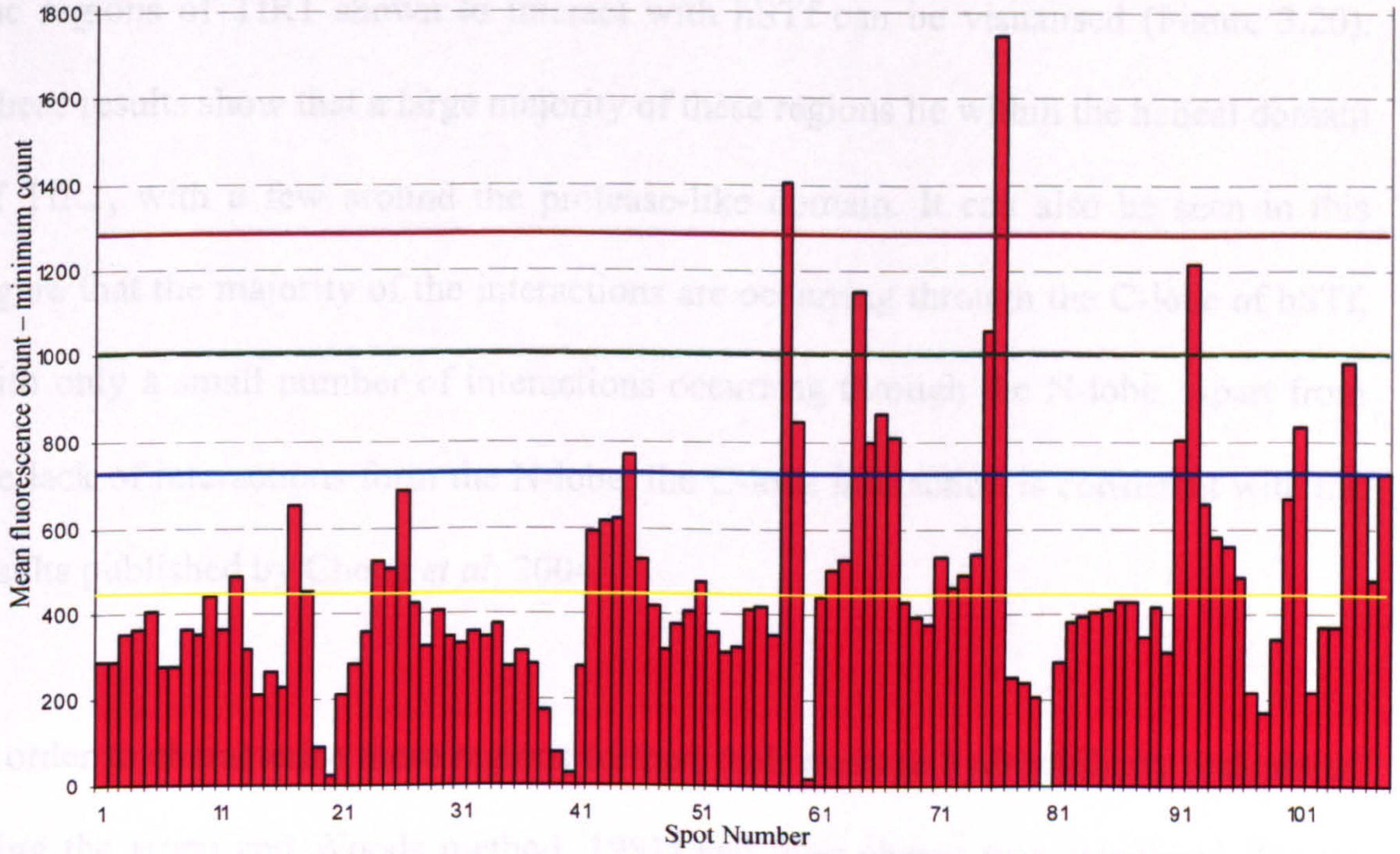


Figure 3.19. TfR1 Peptide library probed with FITC-hSTf. A: Peptide library scanned under UV using a SynGene GelDoc System and quantified using SynGene Genetools. Values represent the mean fluorescence count per spot with the minimum highlighted in yellow. B: Plot of mean fluorescence count minus the minimum spot fluorescence (630.58). Yellow line: Mean. Blue, green and brown line: Mean +1, +2 and +3 standard deviation respectively.

fluorescence of each spot, highlighting the average MFC and also the MFC at +1, +2 and +3 standard deviation (SD) above the average. By determining which spots fluoresce above +1 SD, +2 SD and +3 SD gave an indication of how significant this fluorescence is in terms of hSTf-TfR1 interaction. Taking all spots that have a MFC greater than +1 SD above the average ensures that only the top 16% of all spots, in order of brightness, are chosen. A MFC greater than +2 SD above the average, this percentage is reduced to 2.3% and greater than +3 SD, the percentage is reduced further to 0.15% (Harris *et al*, 2005).

Only 13 spots were detected with a MFC greater than 1 SD above the average. These include spots 46, 58, 59, 64, 65, 66, 67, 75, 76, 91, 92, 101 and 105. By highlighting these areas on the pdb image of the TfR1-hSTf complex (Accession number: 1SUV), the regions of TfR1 shown to interact with hSTf can be visualised (Figure 3.20). These results show that a large majority of these regions lie within the helical domain of TfR1, with a few around the protease-like domain. It can also be seen in this figure that the majority of the interactions are occurring through the C-lobe of hSTf, with only a small number of interactions occurring through the N-lobe. Apart from the lack of interactions form the N-lobe, the C-lobe interaction is consistent with the results published by Cheng *et al*, 2004.

In order to characterise these regions further, their average hydrophilicity (calculated using the Hopp and Woods method, 1981) and their charge was calculated (Figure 3.21a). These characteristics were then mapped onto the receptor image so they could be compared with the characteristics of their corresponding hSTf interacting regions (Figure 3.21b). As expected, the majority of residues on the surface of the

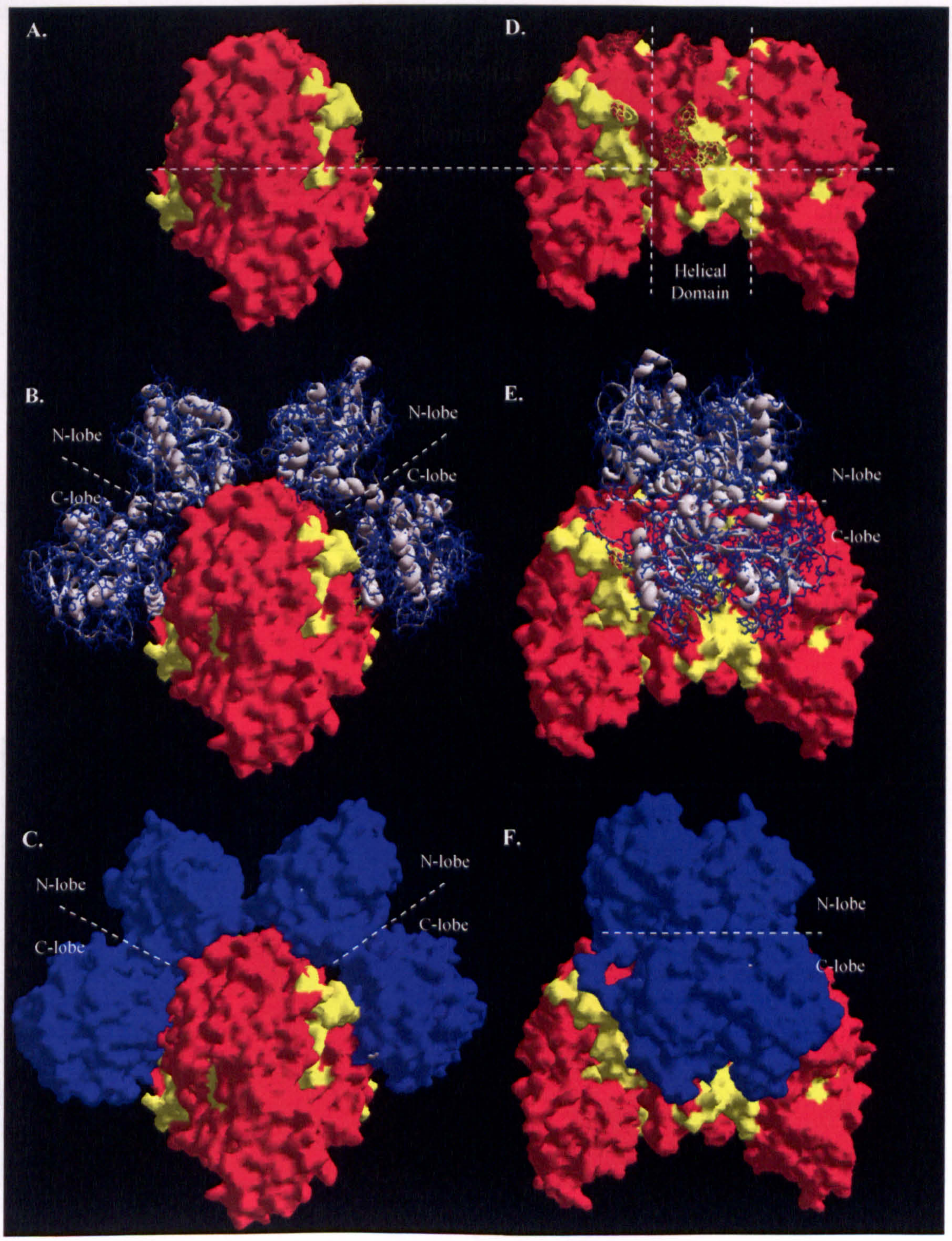


Figure 3.20. Molecular surface images of the TfR1 dimer (A and D) and the hSTf-TfR1 complex (B, C, E and F), highlighting regions shown by peptide library analysis to interact with hSTf. The TfR1 dimer is coloured red with interacting regions in yellow and hSTf coloured blue. Surface generated using Deep Viewer / Swiss PDB Viewer and using PDB Accession Number: 1SUV, Cheng *et al*, 2004.

A.

Peptide No.	Sequence	Positive charges (Arg + Lys)	Negative Charges (Asp + Glu)	Net Charge	Average Hydrophilicity (Hopp and Woods, 1981)
46	FNHTQFPPSRSSG	1	0	+1	-0.1
58	DHYVVVGAQRDAW	1	2	-1	-0.2
59	AQRDAWGPGAAKS	2	1	+1	0.3
64	FQPSRSIIFASWS	1	0	+1	-0.6
65	IFASWSAGDFGSV	0	1	-1	-0.7
66	GDFGSVGATEWLE	0	3	-3	-0.1
67	ATEWLEGYLSSLH	0	2	-2	-0.5
75	GQFLYQDSNWASK	1	1	0	-0.3
76	SNWASKVEKLTLD	2	2	0	0.3
91	DIKEMGLSLQWLY	1	2	-1	-0.4
92	SLQWLYSARGDFF	1	1	0	-0.6
101	FWGSGSHTLPALL	0	0	0	-0.9
105	TLFRNQLALATWT	1	0	+1	-0.8

B.

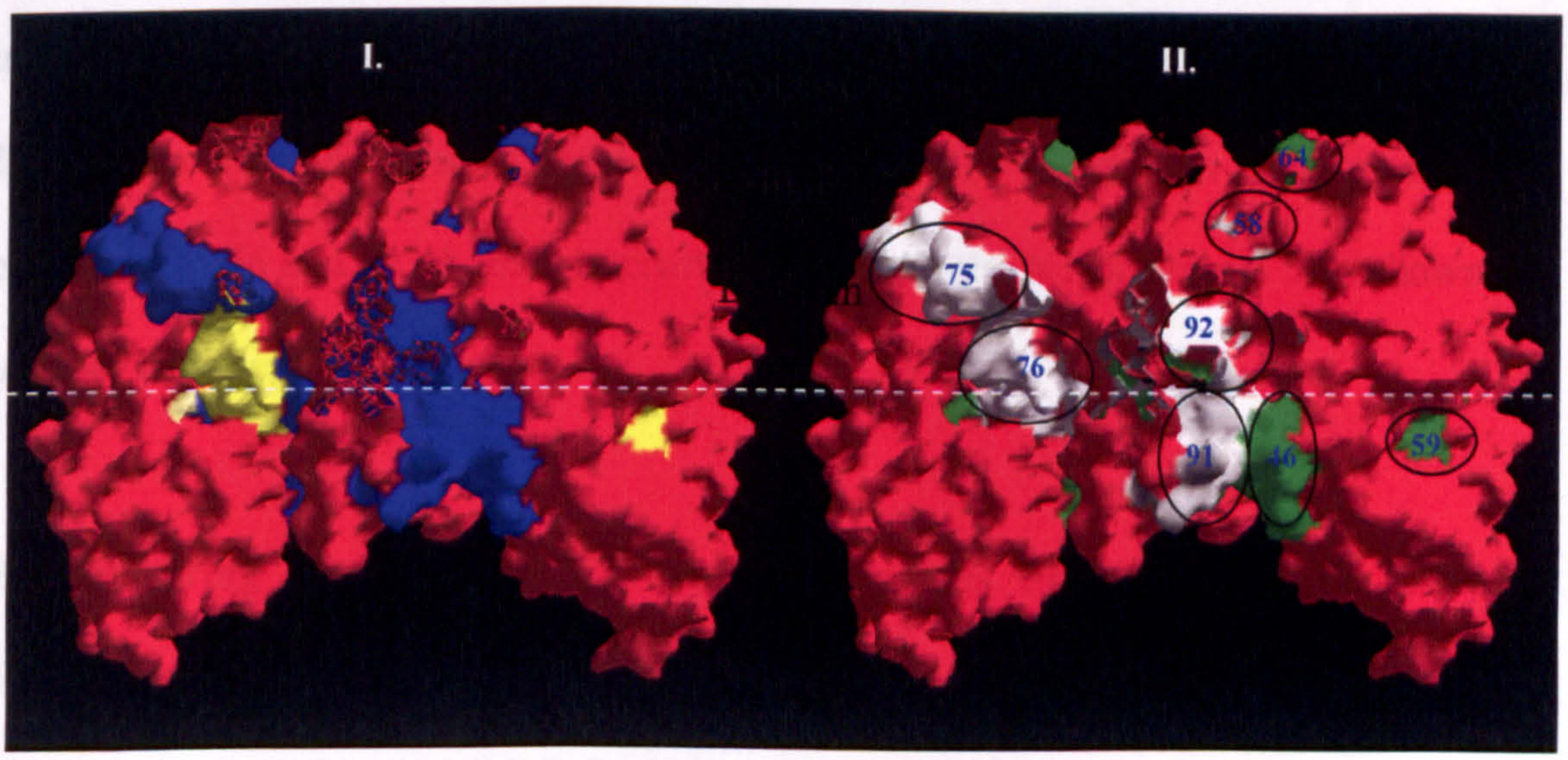


Figure 3.21. Calculated hydrophilicity and charge of TfR1 regions proposed to interact with hSTf (A) and their location on the molecular surface of TfR1 (B). The TfR1 dimer is coloured red with hydrophilic and hydrophobic interacting regions in blue and yellow, respectively (BI) and positive, negative and neutral interacting regions in green, grey and white, respectively (BII). Interacting regions are labelled corresponding to their peptide number (BII). Surface generated using Deep Viewer / Swiss PDB Viewer and using PDB Accession Number: 1SUV, Cheng *et al,* 2004.

protein are hydrophilic. However, two of the regions that are shown to interact are hydrophobic (peptides 59 and 76). These two regions could possibly be involved in hydrophobic interactions with hSTf, especially peptide 76 as it has a MFC greater than +3 SD above the mean. With relation to charge, the majority of interacting regions are neutral with only three regions being positively charged. However, apart from region 64, these positively charged regions have all a MFC below +2 SD above the mean, which suggests that they might not be significant.

By creating a 3D image using Deep Viewer/Swiss PDB Viewer and the PDB file of the TfR1-hSTf complex (Accession Number: 1SUV), the TfR1 dimer and one of the hSTf molecules can be disintegrated away to reveal the possible binding sites on one hSTf (Figure 3.22). Figure 3.22c highlights the 10 hSTf residues which are discussed in section 3.2.2 and that are proposed by Cheng *et al*, 2004 to interact with TfR1. These two main regions in the C1 domain of hSTf are shown to interact with peptides 75, 76 and 92. A third interacting region in the N1 domain is also shown to interact with peptides 58 and 64.

By comparing the proposed TfR1 binding regions of hSTf with the surface characteristics of the hSTf model in section 3.2.3, the characteristics of these specific binding sites can be visualised. From this it can be seen that 8 out of the 10-hSTf residues that are proposed by Cheng *et al*, 2004 to bind TfR1 are hydrophilic, although interspersed with numerous hydrophobic residues. Therefore, one possible interaction could be with the large hydrophobic region (peptide 76) of TfR1 and a number of these hydrophobic residues. Peptide 76 had the highest MFC in these studies, which implies that this interaction with hSTf might be one of the strongest.

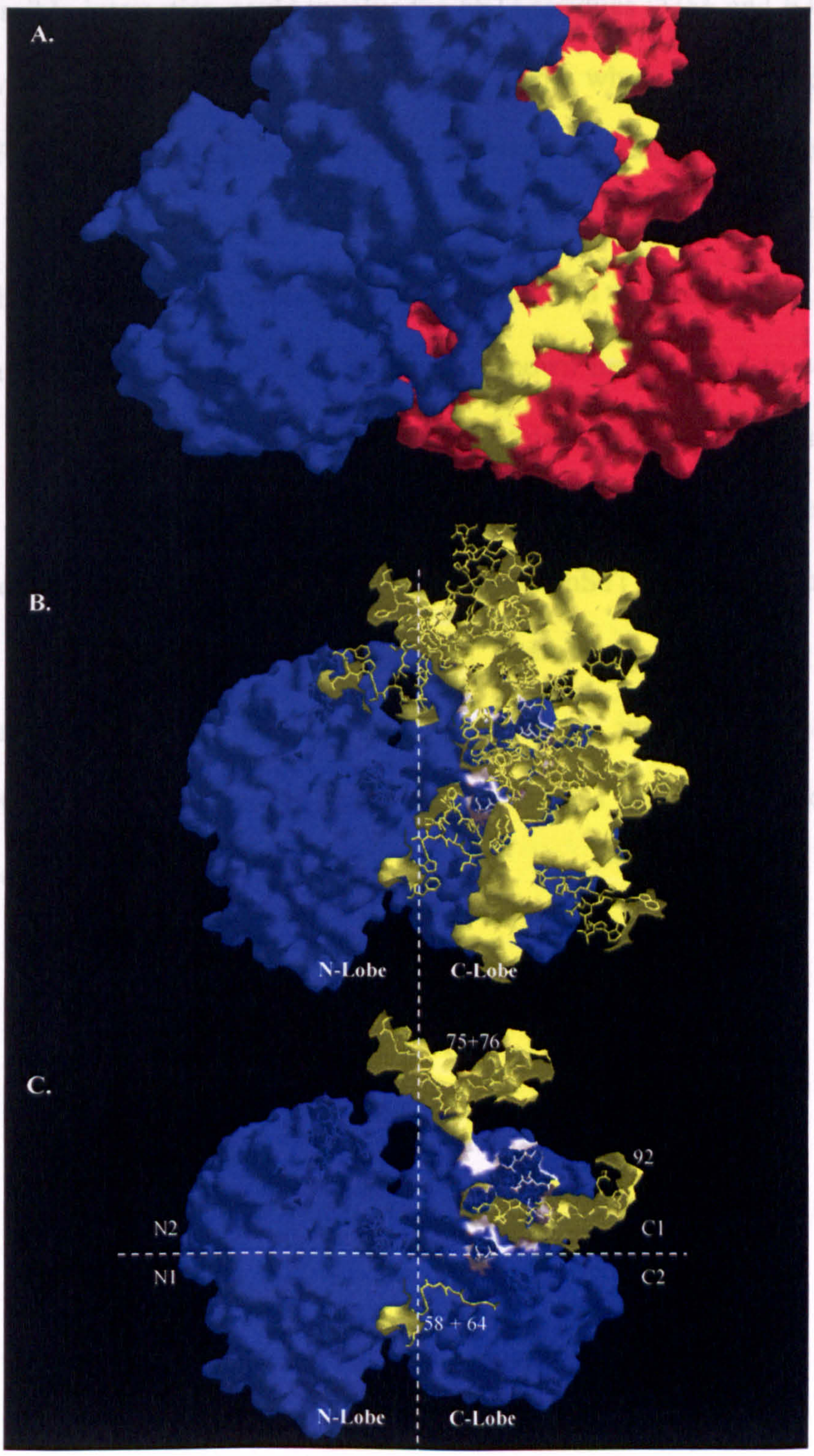


Figure 3.22. Disintegrating molecular images (A→C) of hSTf-TfR1 complex, unveiling the receptor binding regions of hSTf. TfR1 dimer is coloured red with interacting regions in yellow. The transferrin is coloured blue with the published TfR1 binding residues in white. Surface generated using Deep Viewer / Swiss PDB Viewer and using PDB Accession Number: 1SUV, Cheng *et al,* 2004.

Another possible interaction could involve the hSTf residues D356 and E357. These two negatively charged residues are present just below where the TfR1 peptide 92 binds hSTf and in a region that would line up with the positively charged TfR1 peptide 46, following binding of the two molecules. TfR1 peptide 58 and 64 are also shown by these studies to be involved in hSTf binding. As with peptide 76, these two are predominantly hydrophobic, and as they are shown here to bind to a region of hSTf that is also predominantly hydrophobic, this could be the form of this interaction. The high degree of hydrophobicity identified in many of these regions proposed to be involved in the interaction between hSTf and its receptor TfR1, suggests a strong hydrophobic nature of this interaction. Since both molecules present mainly hydrophilic surfaces, such interactions involve regions of both proteins which are not exposed, and therefore implies structural/conformational changes upon interaction.

4. Protein Expression and Mutagenesis

This chapter focuses on the expression of hLf and hSTf using *E.coli* and *S.cerevisiae* cells. There are many advantages and disadvantages for using either of these very different cells for recombinant protein expression. A few reasons for using *E.coli* is that these cells are relatively low maintenance, they have a quick generation time, much is know about them and they are relatively easy to use in cloning. In addition, a large number of different strains are readily available from companies such as Novagen and Promega. *S.cerevisiae* cells are also well known, relatively easy to grow and maintain, and not easily outgrown by contaminants. However, one of the main advantages of using *S.cerevisiae* over *E.coli* is that these are eukaryotic cells and therefore have very similar protein processing pathways as mammalian cells.

4.1. Expression of hSTf and hLf in *Escherichia coli*

4.1.1. Introduction

Three different strains of *E. coli* were used in the following hLf and hSTf expression studies. BL21 (DE3), Origami (DE3) Plys and Origami were all purchased from Novagen (section 2.1.1). Both of these Origami strains have mutations in both the thioredoxin reductase (trxB) and glutathione reductase (gor) genes, which greatly enhances disulfide bond formation in the cytoplasm (Xiong *et al,* 2005).

4.1.2. hLF Expression

The source of hLf cDNA for these studies was in the form of a pUC18 vector, kindly donated by Delta Biotechnology Ltd. Before using this vector in any experiments, an

RFLP study was carried out to check its integrity. This study comprised five different digests: *Bam*HI, *Pst*I, *Hind*III, *Bam*HI / *Pst*I and *Nde*I. These digests were visualised on a 1% (w/v) agarose gel (Figure 4.1). This gel shows the hLf insert to be flanked by two *Bam*HI sites and also two *Hind*III sites. These two sites, including other sites, were confirmed by sequencing the entire hLf cDNA (Cytomyx, Cambridge). The sequence was also used to ensure that no mutations were present in the coding frame. Only two silent mutations were found.

4.1.2.1. Constructs

Two vectors were explored for the expression of hLf in *E. coli*: pET12a and pALTER-Ex1 (section 2.1.7). Both plasmids contain a T7 promoter, a β-lactamase gene for ampicillin resistance, an OriR region for plasmid replication in *E. coli* and their own unique multiple cloning sites. There are also a number of differences: pALTER-Ex1 also contains a gene encoding tetracycline resistance, a *tac* promoter and an SP6 promoter, whereas pET12a contains an *ompT* leader for potential periplasmic export of the target protein. The main reason for wanting to export the protein to the periplasm is that it is thought to improve protein folding and solubility. However, in order to compare protein expression with and without this leader, two separate constructs were made: hLf-pET12a and hLf-modpET12a.

4.1.2.2. Creation of hLf pALTER-Ex1

Since pALTER-Ex1 is a multifunctional vector used for both protein expression and mutagenesis, this was therefore the ideal plasmid for preliminary attempts to express hLf. The hLf cDNA was digested out of pUC18-hLf using *Bam*HI and purified from an agarose gel using a Qiagen Mini Elute Gel Extraction Kit (section 2.1.8)

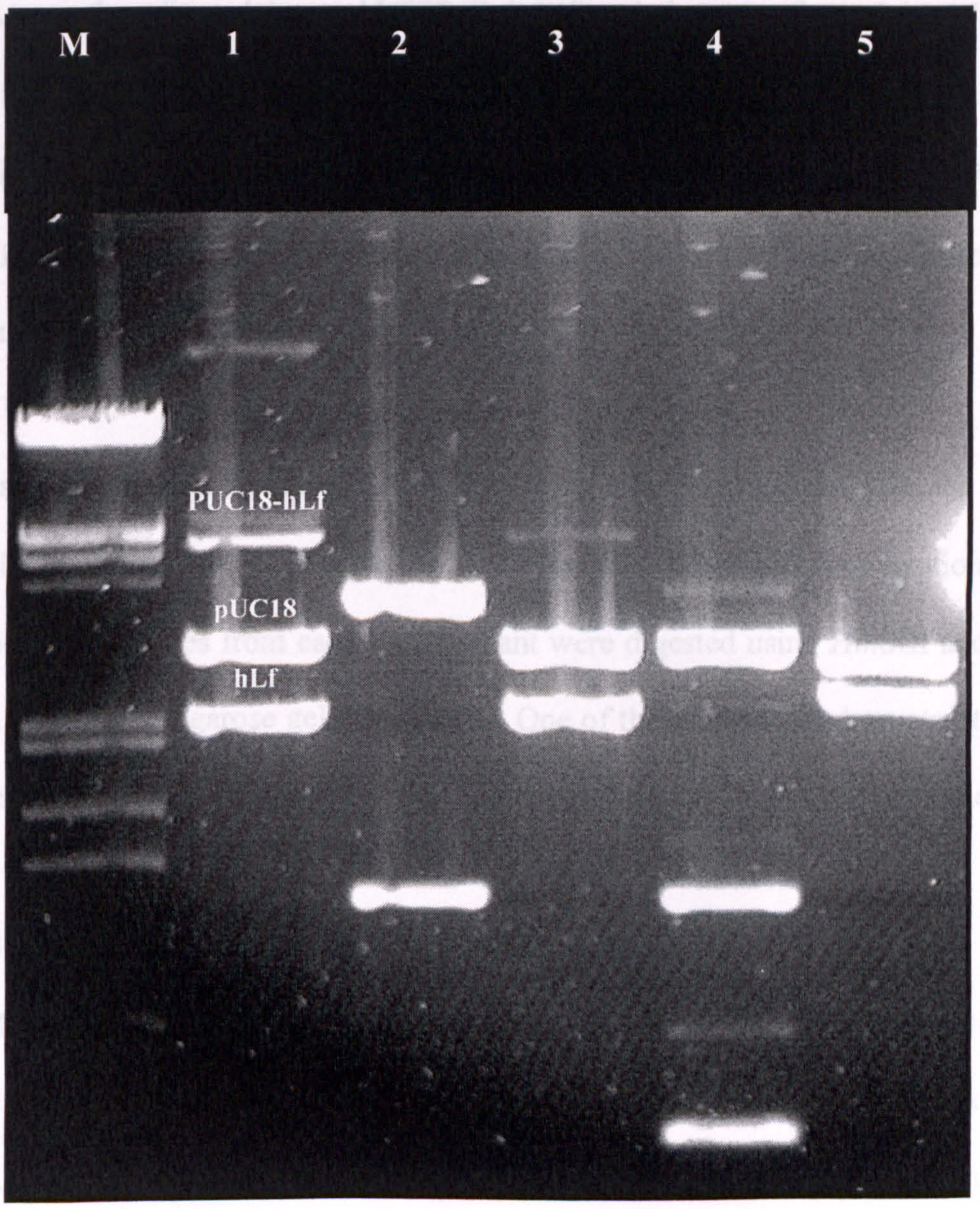


Figure 4.1. An RFLP study of the pUC18-hLf construct on a 1% (w/v) Agarose Gel. M: *Eco*R1 / *Hind*III lambda marker, 1: *Bam*HI, 2: *Pst*I, 3: *Hind*III, 4: *Bam*HI / *Pst*I and 5: *Nde*I.

following electrophoresis. At the same time, pALTER-Ex1 was also digested with *Bam*HI, treated with phosphatase to remove the phosphate group from the 5' end and then purified using a Mini Elute Reaction Cleanup kit (see section 2.1.8). The hLf cDNA was then ligated into pALTER-Ex1 before being transformed into JM109 cells and grown at 37°C overnight on LB agar plates containing 100μg/mL ampicillin. Five transformants were then selected, grown up overnight at 37°C in 5mL LB broth containing 100μg/mL ampicillin, and then plasmids were then extracted using a Spin Mini Prep kit (section 2.1.8).

Because hLf was ligated into pALTER-Ex1 using a single restriction site, the insert could be orientated in two different ways (Figure 4.2). To check its orientation, 2μL plasmid prep samples from each transformant were digested using *Hind*III and then run on a 1% (w/v) agarose gel (Figure 4.3). One of the T7 orientated constructs was selected for expression studies in *E. coli* (DE3) carrying the T7 polymerase gene.

4.1.2.3. Creation of hLf-pET12a (with OmpT leader)

The hLf cDNA from the hLF-pUC18 vector was cloned into the *Bam*HI site of pET12a using the same procedure as for the ligation of hLf into pALTER-Ex1. As before, twelve transformants were selected, grown up overnight at 37°C in 5mL LB broths containing 100μg/mL ampicillin, and their plasmids were then extracted using a Spin Mini Prep kit.

Because hLf was ligated into pET12a using a single restriction site, the insert can be present in two orientations (Figure 4.4). The hLf orientation for all twelve

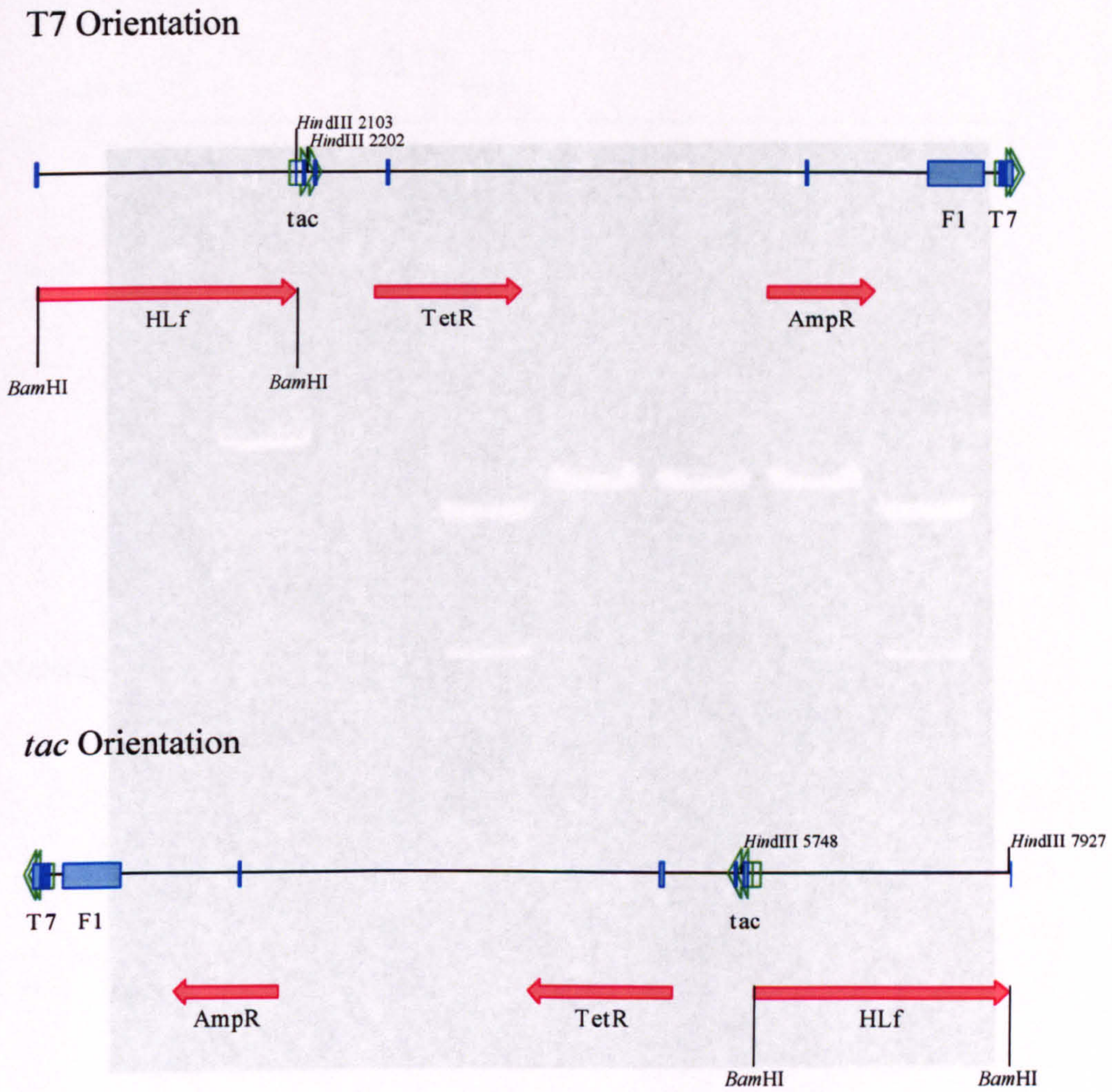


Figure 4.2. Linear vector maps of the two possible ligation products following the cloning of hLf into pALTER-Ex1 using the *Bam*HI restriction site. The two orientations can be visualised following digestion with *Hind*III. T7 ori: 7851bp and 99bp, tac ori: 5771bp and 2179bp.

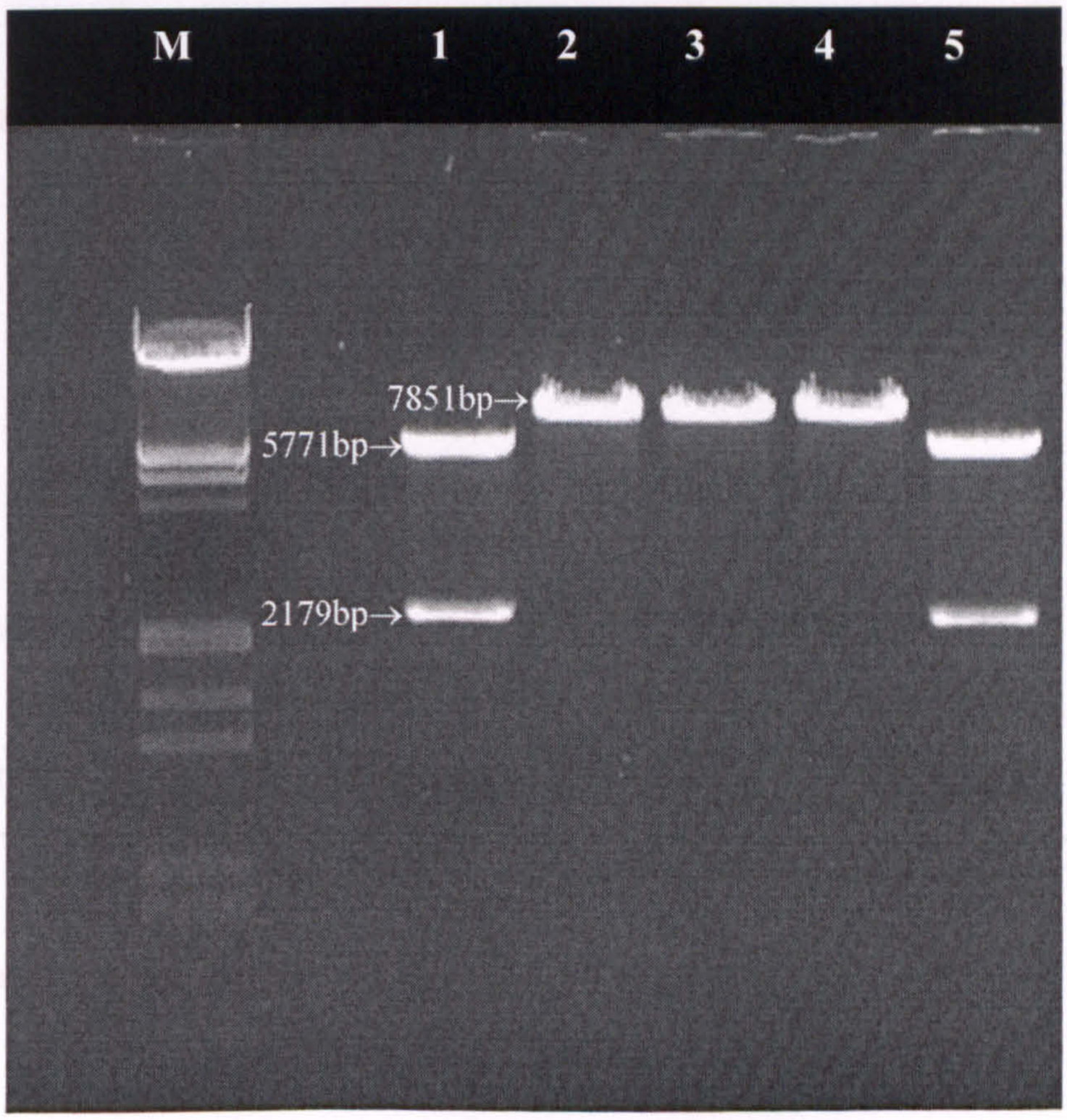
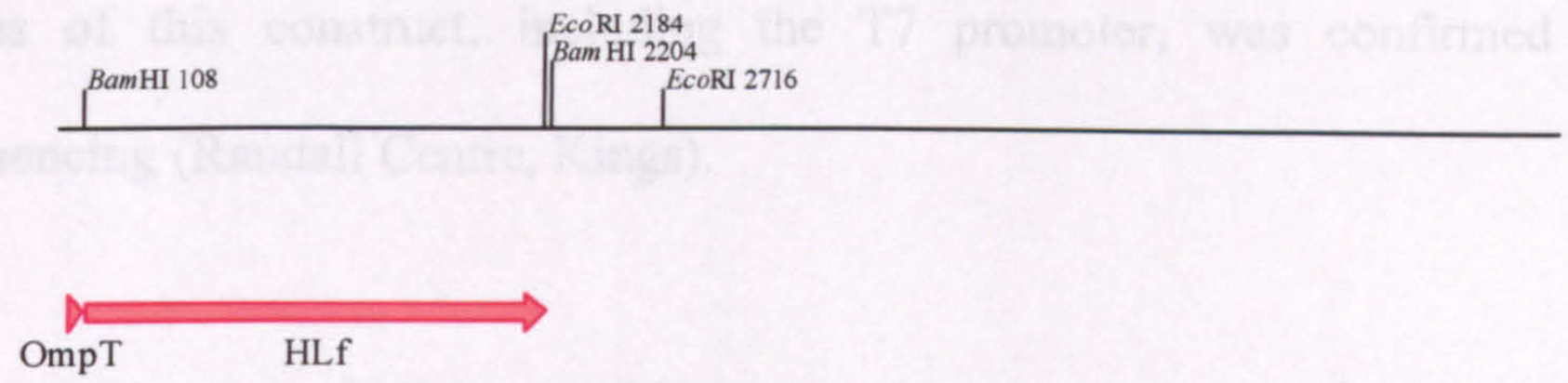


Figure 4.3. Agarose Gel Electrophoresis showing the orientation of the hLf cDNA in pALTER-Ex1 following digestion of five transformants with *Hind*III. M: *Eco*R1 / *Hind*III lambda marker and 1-5: transformants.

Orientation 1



Orientation 2

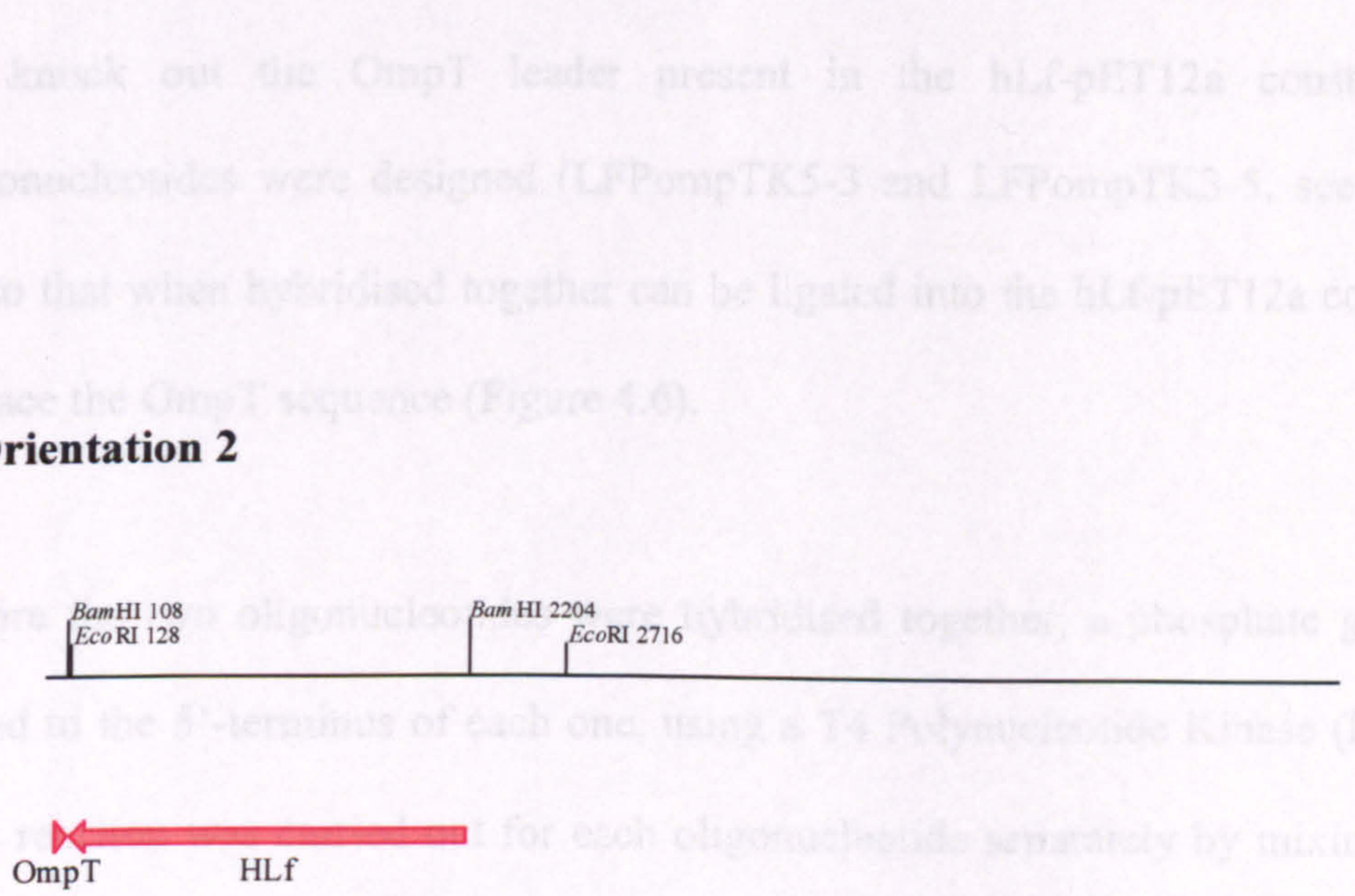


Figure 4.4. Linear vector maps of the two possible ligation products following the cloning of hLf into pET12a using the *Bam*HI restriction site. The two orientations can be visualised following digestion with *Eco*R1. The correct orientation for expression using the T7 promoter is Ori1. Ori1: 6238bp and 532bp, Ori2: 4182bp and 2588bp.

transformants was checked by observing the fragment sizes produced following 1% (w/v) agarose gel electrophoresis of *Bam*HI and *Eco*RI hLf-pET12a plasmid digests (Figure 4.5). Out of the twelve, only transformant eleven carried the hLf-pET12a plasmid with the hLf cDNA in the correct orientation. The sequence of the first 500 bases of this construct, including the T7 promoter, was confirmed by DNA sequencing (Randall Centre, Kings).

4.1.2.4. Creation of hLf-modpET12a (without OmpT leader)

To knock out the OmpT leader present in the hLf-pET12a construct, two oligonucleotides were designed (LFPompTK5-3 and LFPompTK3-5, see appendix II) so that when hybridised together can be ligated into the hLf-pET12a construct to replace the OmpT sequence (Figure 4.6).

Before the two oligonucleotides were hybridised together, a phosphate group was added to the 5'-terminus of each one, using a T4 Polynucleotide Kinase (Promega). This reaction was carried out for each oligonucleotide separately by mixing 2μL of 100μM oligonucleotide, 2.5μL of 10x Kinase buffer, 2.5μL of 10nM ATP, 1μL of T4 Polynucleotide Kinase with 17μL of MilliQ water. The mixture was then incubated at 37°C for 30 minutes, then heated at 70°C for 10 minutes to deactivate the enzyme. These two 25μL reaction mixtures were then mixed together and heated to 75°C for 5 minutes and then cooled at 1°C/min to allow the two oligonucleotides to slowly hybridise.

Following hybridisation of the two oligonucleotides, hLf-pET12a was digested with *Xba*I and *Sal*I to excise the OmpT leader. The digested plasmid was then purified by

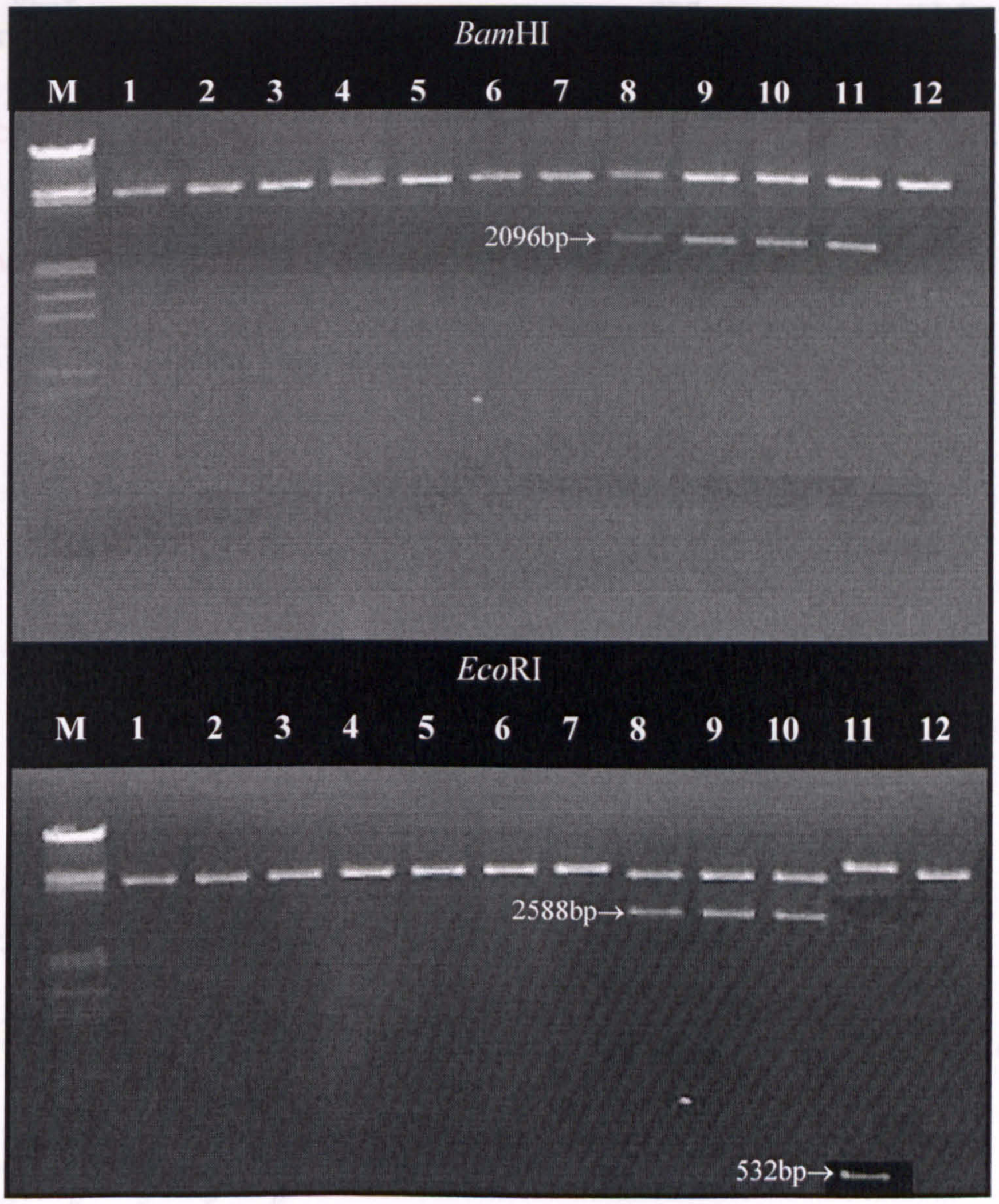
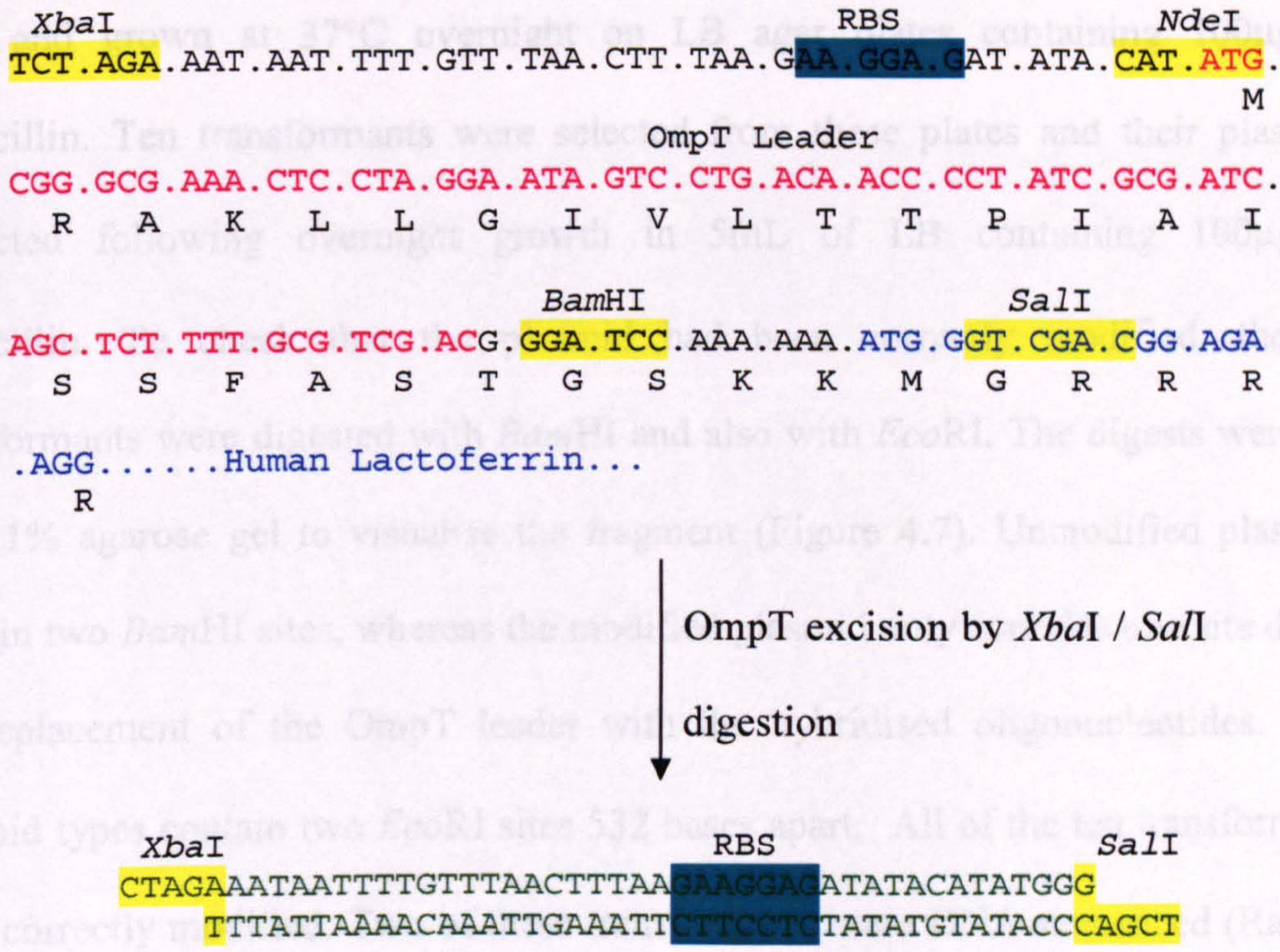


Figure 4.5. Agarose Gel Electrophoresis showing the orientation of the hLf cDNA in pET12a following digestion of five transformants with *Bam*HI *and Eco*RI. M: *Eco*R1 / *Hind*III lambda marker, 1-12: Transformants

hLf-pET12a Construct



hLf-modpET12a Constuct

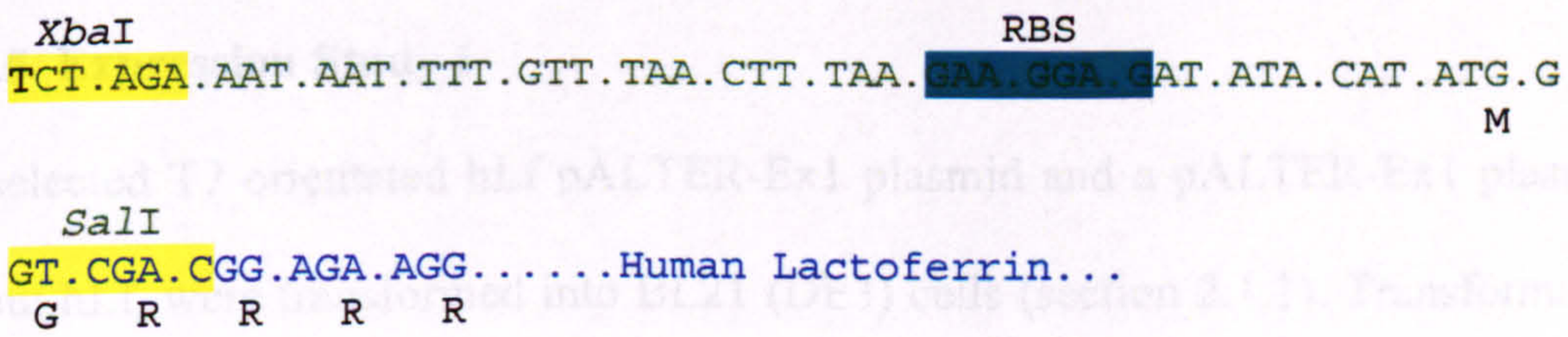


Figure 4.6. Creation of the hLf-modpET12a construct by replacing the OmpT leader sequence with a 50 base long double-stranded oligonucleotide. Sequence in black: pET12a, red: OmpT, blue: hLf insert and green: the hybridised LFPompTK5-3 and LFPompTK3-5 insert. The lactoferrin isoform used in these studies included four arginines after residue 1 (glycine), unlike the three arginines observed in other isoforms.

agarose gel electrophoresis and gel extraction before being ligated with an equal molarity of hybridised oligonucleotides. This ligation was transformed into JM109 cells and grown at 37°C overnight on LB agar plates containing 100µg/mL ampicillin. Ten transformants were selected from these plates and their plasmids extracted following overnight growth in 5mL of LB containing 100µg/mL ampicillin. To check that the plasmid had been correctly modified, the ten transformants were digested with *Bam*HI and also with *Eco*RI. The digests were run on a 1% agarose gel to visualise the fragment (Figure 4.7). Unmodified plasmids contain two *Bam*HI sites, whereas the modified plasmid only contains one site due to the replacement of the OmpT leader with the hybridised oligonucleotides. Both plasmid types contain two *Eco*RI sites 532 bases apart. All of the ten transformants were correctly modified. Two of these transformants were DNA sequenced (Randall Centre, Kings) around the modification site to check sequence integrity. Both were correct, but only one transformant was selected by random.

4.1.2.5. Expression Study 1

The selected T7 orientated hLf pALTER-Ex1 plasmid and a pALTER-Ex1 plasmid without hLf, were transformed into BL21 (DE3) cells (section 2.1.1). Transformants were selected for on LB agar plates containing 10µg/mL tetracycline. From these plates, one transformant from each of the two clones were selected and used to inoculate 5mL LB broth (10µg/mL tetracycline) and incubated overnight at 37°C. The following day, 6 x 5mL LB broth were inoculated with 50µL of hLf pALTER-Ex1 overnight culture and 1 x 5mL LB broth was inoculated with 50µL of pALTER-Ex1 overnight culture. After incubation for 3 hours at 37°C on a shaking platform, IPTG was added to three of the hLf constructs to a final concentration of

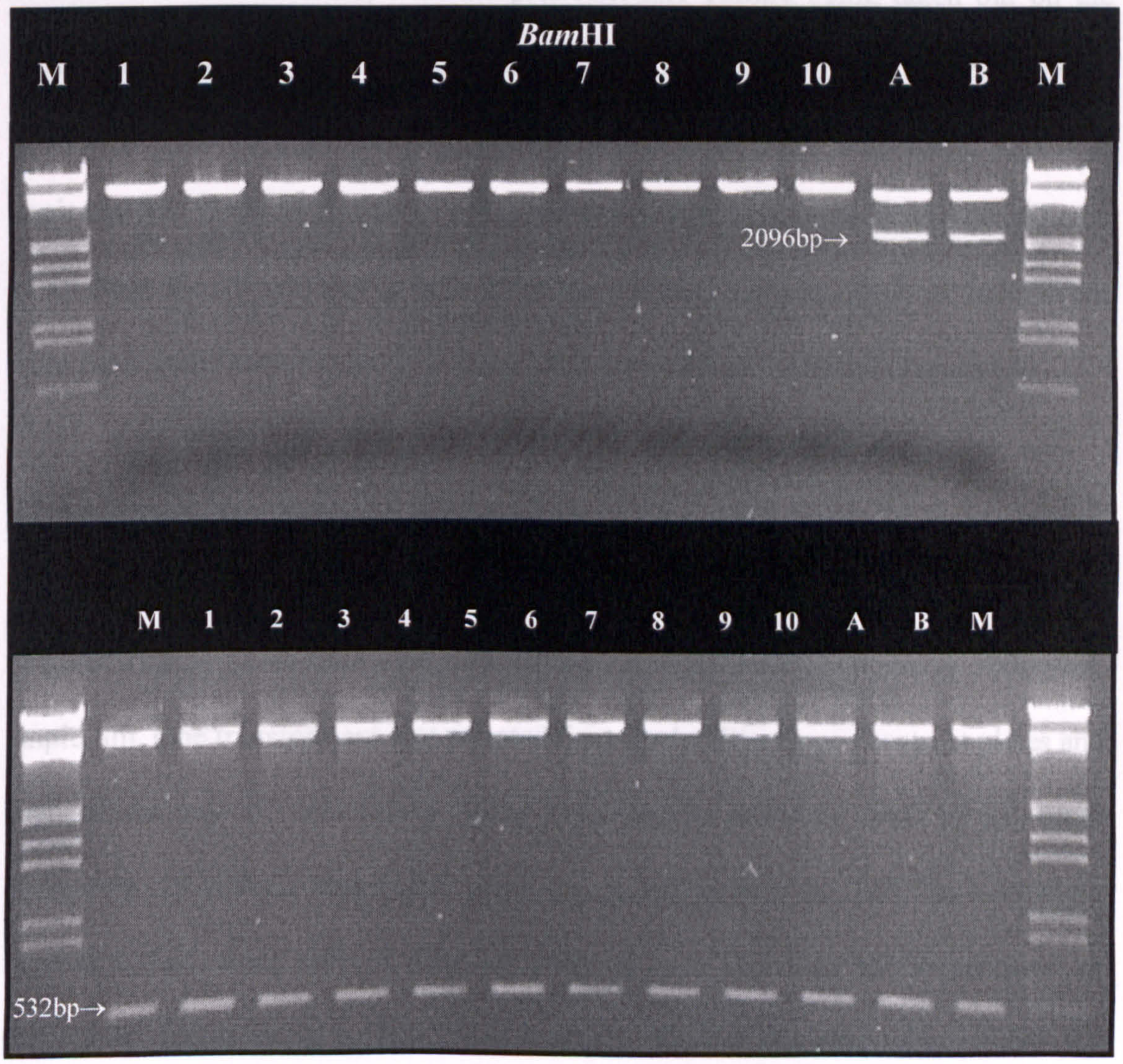


Figure 4.7. Agarose Gel Electrophoresis of 10 hLf-modpET12a transformants following *Bam*HI and *Eco*RI digestion. M: *Eco*R1 / *Hind*III lambda marker, 1-10: hLf-modpET12a transformants and A - B: hLf-pET12a.

1mM. All seven cultures were then placed back in the incubator. After a further two hours, two of the hLf construct cultures were taken out and centrifuged at 2000 *g* for 10 minutes. The supernatant was poured off and the pellet frozen. This was repeated at three and at four hours, with the pALTER-Ex1 culture being taken out on the fourth hour. Frozen pellets were then defrosted at room temperature before being resuspended in 1mL of Cell Suspension buffer. Each 1mL suspension was then sonicated for 30 seconds using a Mullard Ultrasonic Powerunit and then centrifuged at 10,000*g* for 20 minutes to spin down any debris. Samples (5µl) of supernatant from each sample were run on a reducing 8% (w/v) PAGE SDS gel (Figure 4.8).

4.1.2.6. Expression Study 2

hLf-pET12a, hLf-modpET12a and pET12a were transformed into BL21 (DE3) and Origami (DE3) pLys cells and selected on LB agar plates containing 100µg/mL ampicillin. One transformant was selected from each of the three construct plates and each of the two *E. coli* strains plates and used to inoculate 5mL of LB broth containing 100µg/mL ampicillin. After overnight growth at 37°C in a shaking incubator, 1mL from each of these six cultures were tested for the presence of its expression vector using a Spin Mini Prep kit, leaving the remaining culture for expression studies. Once all had been confirmed to contain an expression vector, three different inductions were set up for each of the six cultures (Table 4.1 number 1-18). Each induction experiment was carried out using 5mL of LB broth containing 100µg/mL ampicillin and 100µl of the specific inoculum in a temperature-controlled shaking incubator. Cells were induced by the addition of IPTG to a concentration of 1mM. After induction, cells were harvested, sonicated and their protein expression

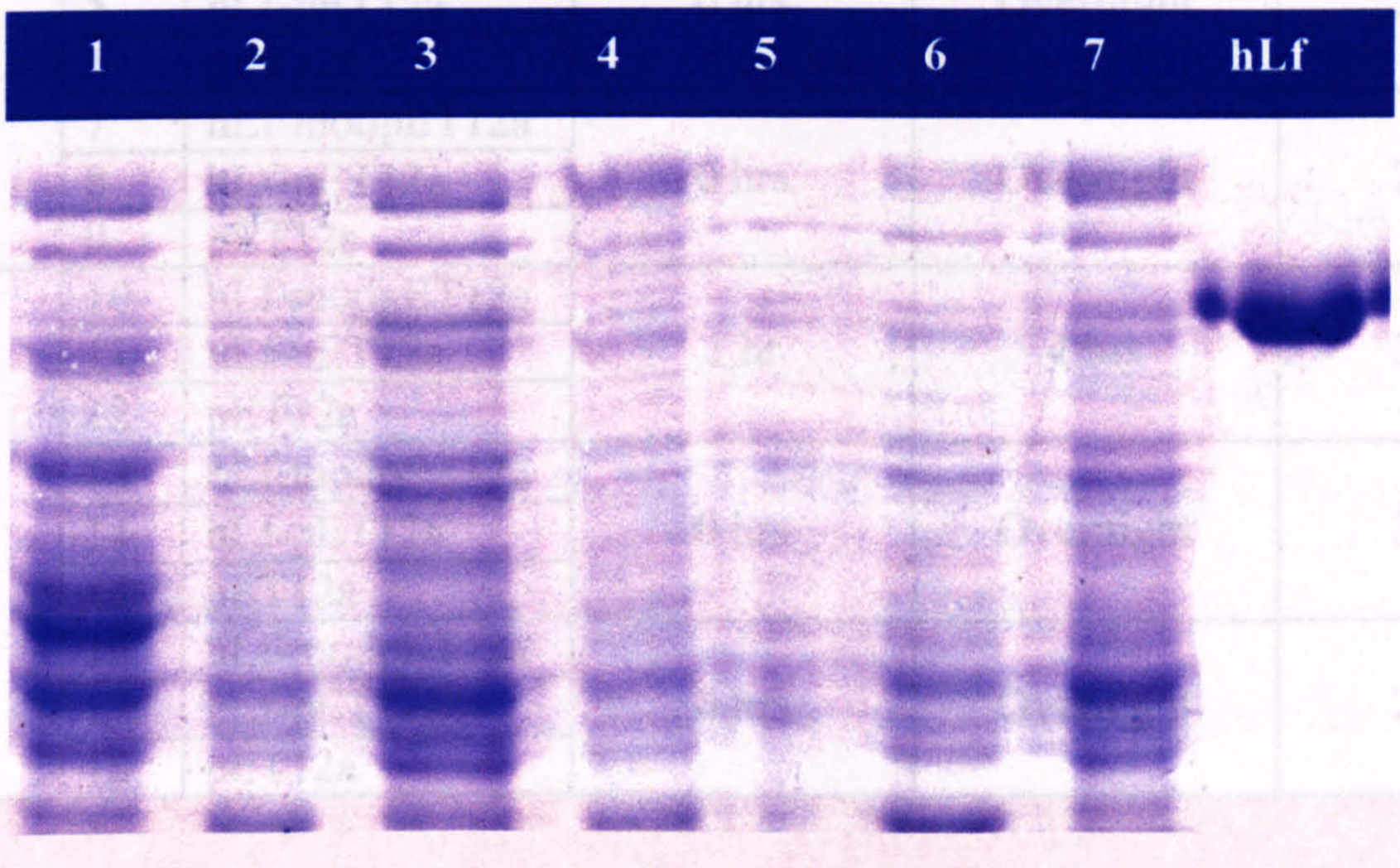


Figure 4.8. Reducing SDS-PAGE of IPTG induction experiments using BL21 (DE3) cells containing hLf pALTER-Ex1. hLf was used as a reference band in order to locate any *E. coli* expression of hLf. 1: 2 hrs non-induced, 2: 2 hrs induced, 3: 3 hrs non-induced, 4: 3 hrs induced, 5: 4 hrs non-induced, 6: 4 hrs induced and 7: pALTER-Ex1 negative control. All samples run were extracted from their soluble fractions.

E. coli strain	No.	Expression Construct	Pre-Induction Growth	Post-Induction Growth	Incubation Temp (°C)
Expression Study 2: IPTG Induction					
Origami (DE3)	1	hLf-modpET12a	1 hr	4 hrs	37
	2	hLf-pET12a			
	3	pET12a			
	4	hLf-modpET12a	0 hrs	Overnight	37
	5	hLf-pET12a			
	6	pET12a			
	7	hLf-modpET12a	0 hrs	Overnight	25
	8	hLf-pET12a			
	9	pET12a			
BL21 (DE3)	10	hLf-modpET12a	1 hr	4 hrs	37
	11	hLf-pET12a			
	12	pET12a			
	13	hLf-modpET12a	0 hrs	Overnight	37
	14	hLf-pET12a			
	15	pET12a			
	16	hLf-modpET12a	0 hrs	Overnight	25
	17	hLf-pET12a			
	18	pET12a			
Expression Study 3: IPTG Induction					
Origami (DE3)	19	hLf-modpET12a	3 hrs	2 hrs	37
	20		4 hrs		
	21		5 hrs		
	22	hLf-pET12a	3 hrs		
	23		4 hrs		
	24		5 hrs		
	25	pET12a	5 hrs		
Expression Study 4: Bacteriophage λCE6 Induction					
Origami	26	hLf-modpET12a	2 hrs	3 hrs	37
	27	hLf-pET12a			
	28	pET12a			
	29	hLf-modpET12a	2 hrs	5 hrs	
	30	hLf-pET12a			
	31	pET12a			

Table 4.1. Induction plan for hLf expression using pET12a in *E. coli* Origami (DE3) pLys, BL21 (DE3) and Origami cells.

profile visualised on an SDS-PAGE (Figure 4.9 numbers 1-18) using the same method as for hLf pALTER-Ex1 induction experiments.

4.1.2.7. Expression Study 3

Using the same method, three more inductions were set up (Table 4.1 number 19-25). Once again, protein expression profiles were visualised using SDS-PAGE (Figure 4.10 numbers 19-25).

4.1.2.8. Expression Study 4: λCE6 Bacteriophage Induction

DE3 deficient cells containing the hLf-modpET12a, hLf-pET12a and pET12a were grown in duplicate (Table 4.1 numbers 26-31) in LB broth containing 0.2% maltose and the appropriate antibiotic. When the OD_{600} of each culture reached between 0.6 and 1 (after approximately 2 hours), $MgSO_4$ was added to a final concentration of 10mM and the λCE6 stock (Novagen) was added to a final concentration of 2-4 x 10^9 pfu/mL. One set of cells was allowed to grow for 3 hours at 37°C, whilst the other set was allowed to grow for a further 2 hours at 37°C. Cells were harvested, sonicated and their protein expression profile visualised on an SDS-PAGE (Figure 4.10 numbers 26-31) using the same method as before.

4.1.2.9. Dot-Blots

Using the method described in section 2.4, four dot-blots were set up using samples 1-31 (Table 4.1). Drops (3μl) of each sample, including two positive controls and a negative control, were used in this study. These dot-blots were then probed with FITC-conjugated Anti-hLf IgG (section 2.4) to look for hLf expression (Figure 4.11).

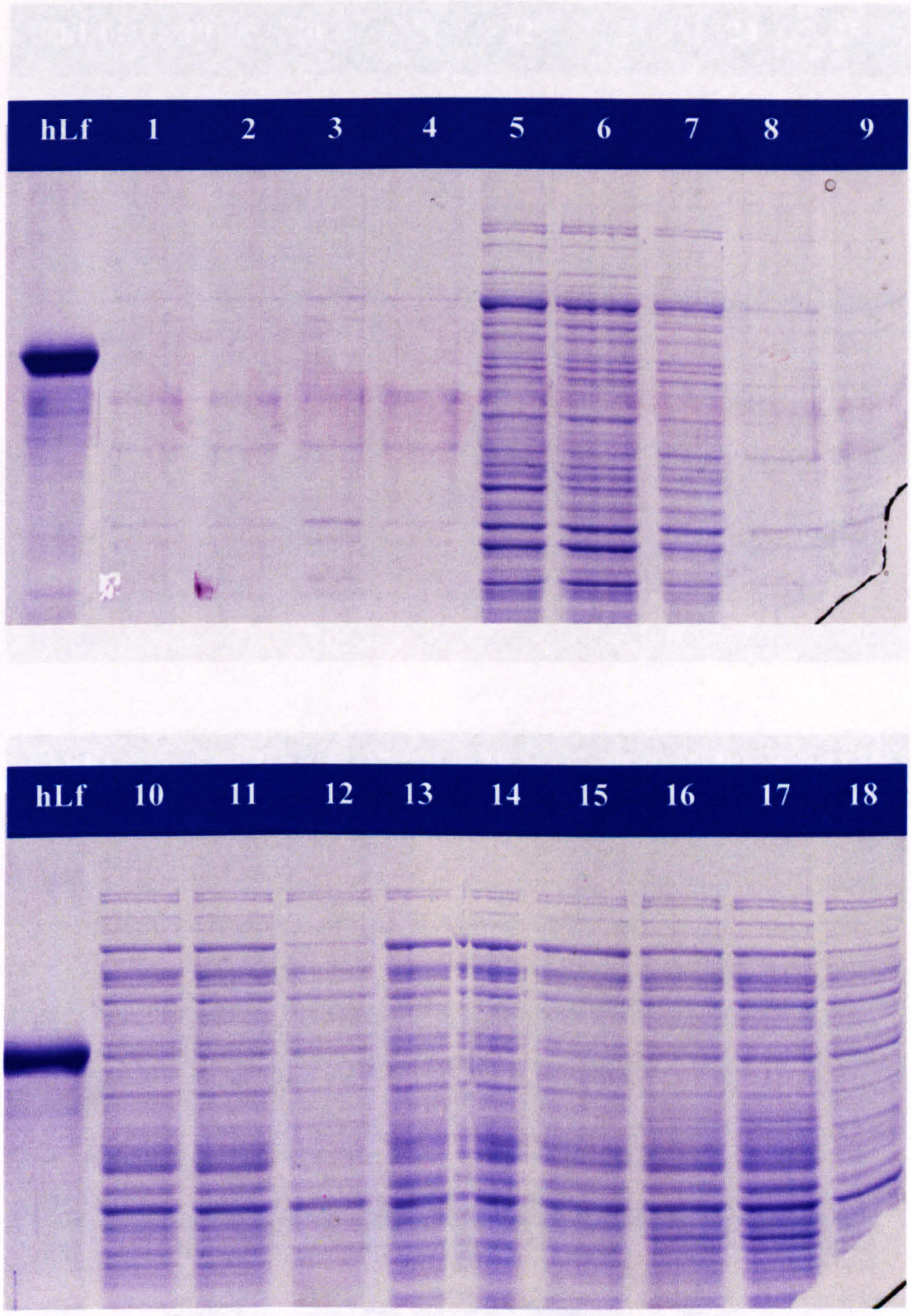


Figure 4.9. Reducing SDS-PAGE of IPTG induction experiments, numbers 1-18. hLf was used as a reference band in order to locate any *E. coli* expression of hLf. Numbers 1-9: Origami (DE3) plys, 10-18: BL2 (DE3) (see table 4.1). All samples run were extracted from their soluble fractions.

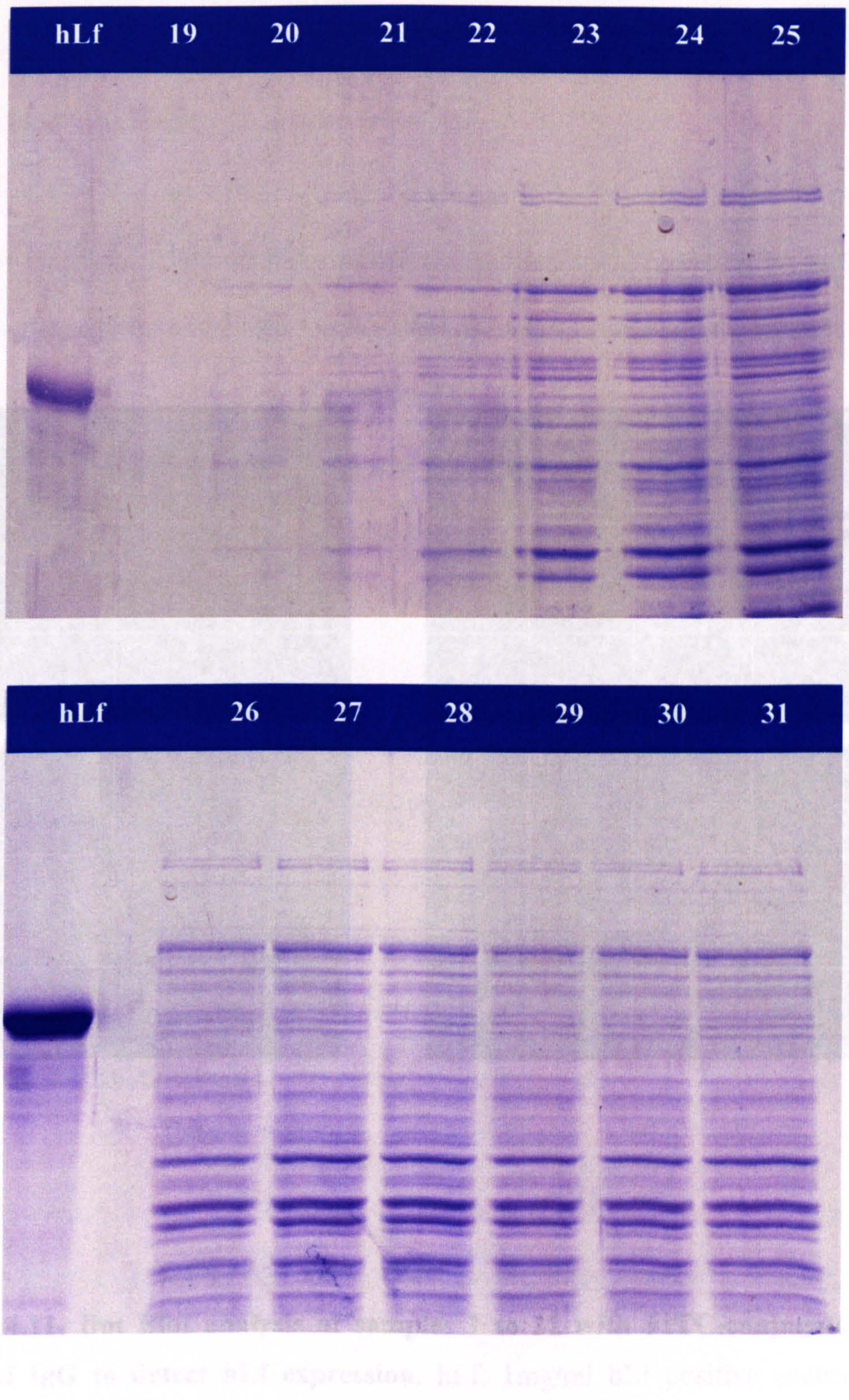


Figure 4.10. Reducing SDS- PAGE of hSTf expression studies 3 and 4, experiment numbers 19-31. hLf was used as a reference band in order to locate any *E. coli* expression of hLf. Numbers 19-25: Origami (DE3) plys IPTG induction, 10-18: Origami (DE3) λCE6 induction (table 4.1). All samples run were extracted from their soluble fractions.

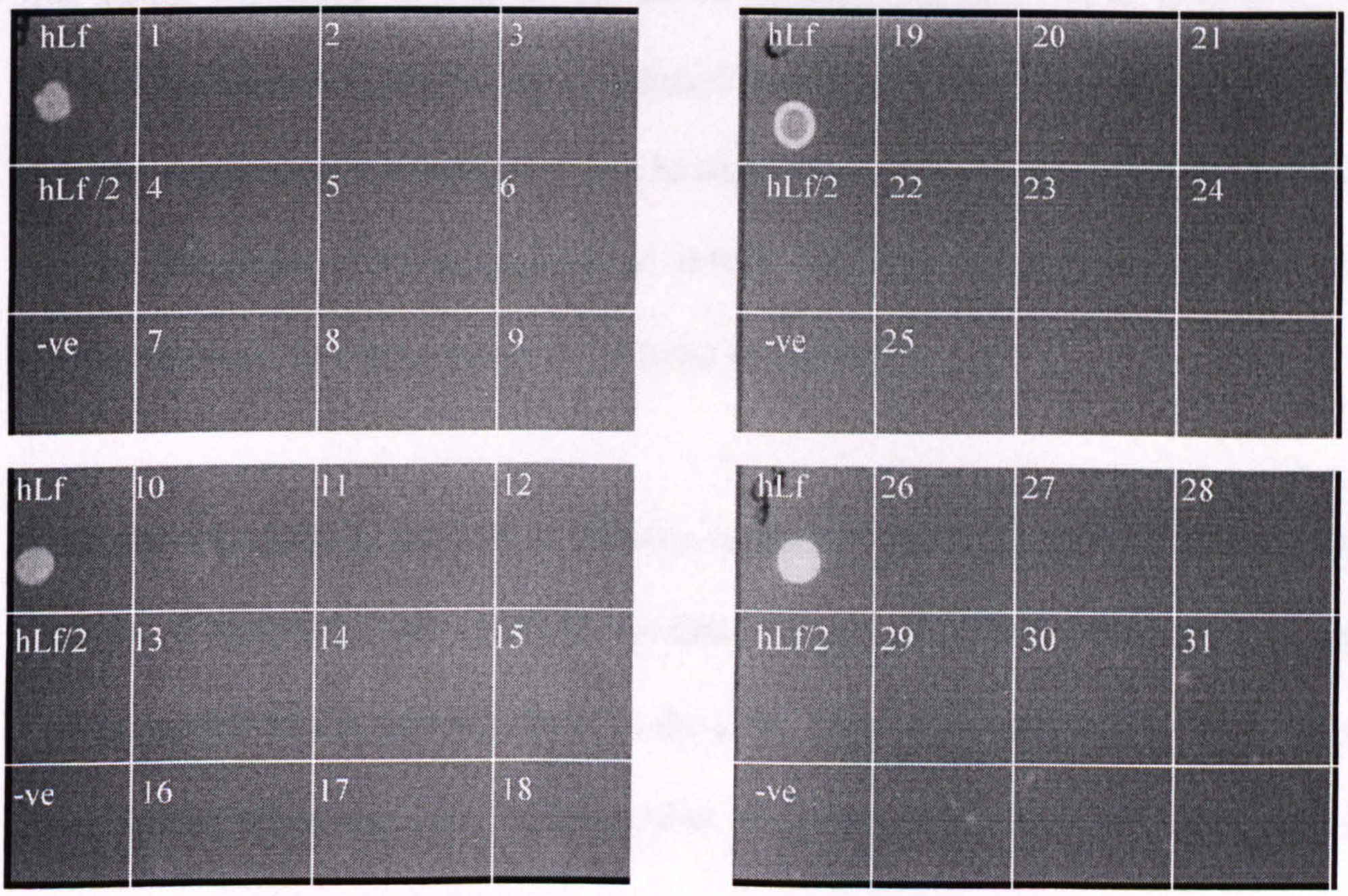


Figure 4.11. Dot Blot analysis of samples 1 to 31 with FITC-conjugated anti-hLf IgG to detect hLf expression. hLf: 1mg/ml hLf positive control, hLf/2: 0.5mg/ml hLf positive control, -ve: water and 1-31: samples from expression studies 2,3 and 4 (table 4.1).

4.1.2.10. Discussion

As described earlier, three expression constructs and four different expression studies were used to try and express hLf in *E. coli*.

In expression study 1, the protein profiles shown on the SDS gel are the same for each of the six cultures and also for the negative control. This suggests that no hLf was being expressed in any of the seven cultures. However, cell growth in the induced cultures was significantly reduced when comparing with the non-induced cultures, suggesting that induction was having some effect on the cells. This could be due to the cells expressing hLf at levels too low to be detected by SDS electrophoresis, but high enough to be toxic to the cells.

In expression study 2, no hLf expression could be detected in any of the eighteen cultures. A reason for this could be the same reason used for the hLf pALTER-Ex1 studies, in that hLf is actually toxic to the cells. Even when induced at 25°C, when cells should be under less pressure due to a reduced metabolic rate, no hLf expression could be observed.

To try and overcome this possible selective pressure that causes cells not to express hLf, pre-induction growth in expression study three was lengthened to allow the cell mass to reach greater levels in order to increase the chance that protein will be expressed following induction. As before, no extra bands corresponding to hLf were visible. What could be noted however, was hLf-modpET12a cells grew much slower than that for hLf-pET12a. This could be due to hLf-modpET12a expressing hLf at

higher levels than hLf-pET12a due to the lack of the fusion leader. However, as expression was not detected in either of the cells, this is not conclusive.

A problem with transforming T7 controlled expression vectors into DE3 lysogenic hosts is that; even without IPTG induction there will always be low-level expression or leak of the target protein. Therefore, if the target protein is toxic to the cell, there will always be a selective pressure against expressing it. This could be a reason for the lack of hLf expression in expression studies one to three.

In expression study 4, the bacteriophage induction system from Novagen was employed. Bacteriophage λCE6 is a recombinant lambda phage containing the T7 RNA polymerase gene. The polymerase gene is cloned into the *int* gene such that it is constitutively transcribed from the p_L and p_I promoters of the phage during infection. Using this, it is therefore possible to selectively induce cells containing the T7 expression construct during exponential phase by infecting them with the bacteriophage. Unfortunately though, even using this induction system, no hLf expression could be detected.

As protein expression may not be easily detected using SDS-PAGE, especially when at low levels, all 31 samples, from expression studies two to four, were blotted onto a nitrocellulose membrane and then probed with FITC-conjugated anti-hLf IgG. None of the 31 samples reacted with the FITC-conjugated anti-hLf IgG. As discussed in Chapter 7, these results, taken together with the SDS-PAGE results, suggest that no hLf was expressed in any of these four studies.

4.1.3. hSTf Expression

The source of hSTf cDNA for these studies was in the form of a pcG13 vector, kindly donated by Christopher Joannou (King's College London). The presence of the hSTf cDNA in the plasmid was confirmed by sequencing either ends of the clone using hTrfFwd and hTrfBck oligonucleotides (appendix II). Using these oligonucleotides, 400 bases at either end of hSTf were sequenced, thus confirming its presence.

4.1.3.1. Constructs

Two vectors were used for the cloning and expression of hSTf in *E. coli*. The first one, pALTER-max (section 2.1.7), is a mutagenesis vector that can be used to express target proteins in mammalian cells as well as being able to be replicated in *E. coli*. The second vector, pET21d, is designed for high-level prokaryotic expression of proteins using a T7 promoter (section 2.1.7).

4.1.3.2. Creation of hSTf pALTERmax

In order to clone the hSTf cDNA from the pcG13 vector into pALTER-max, hSTf was PCR amplified (section 2.2.5) using *Pfu* DNA polymerase for high-fidelity DNA synthesis and the same two oligonucleotides used to sequence its ends. These two oligonucleotides, hTrfFwd and hTrfBck, are designed to complement either end of the hSTf cDNA, with overhanging sequence that will create two new restriction sites following PCR amplification (Figure 4.12). Extra sequence was added to the ends of the oligonucleotides to optimise cleavage of these two new sites. This is recommended by NewEngland Biolabs for restriction sites close to the end of DNA fragments.

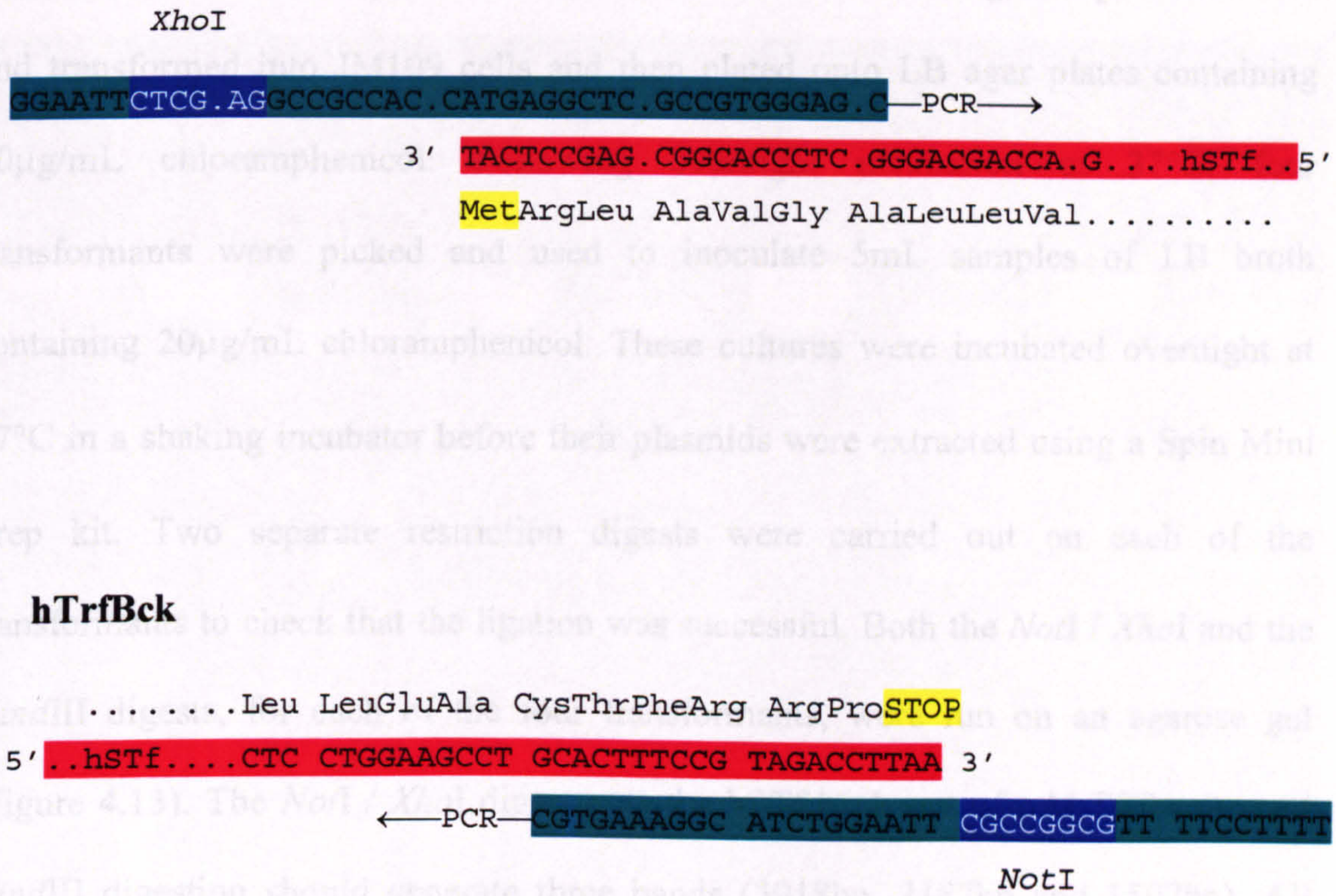


Figure 4.12. A Diagram to show the binding of hTrfFwd and hTrfBck oligonucleotides to hSTf and the creation of two restriction sites at each end during PCR amplification. The sequence highlighted in red represents hSTf, whilst the sequence highlighted in turquoise represents the oligonucleotides. The start codon and stop codon are highlighted in yellow and the two new restriction sites in blue.

The amplified hSTf was purified from the PCR reaction mix by agarose gel electrophoresis and gel extraction before digestion with *Not*I and *Xho*I. At the same time, pALTER-max was digested using the same two enzymes. Both digestions were run on an agarose gel to purify hSTf and pALTER-max from their reaction mixture by gel extraction. The digested hSTf was then ligated into the digested pALTER-max and transformed into JM109 cells and then plated onto LB agar plates containing 20μg/mL chloramphenicol. Following overnight incubation at 37°C, four transformants were picked and used to inoculate 5mL samples of LB broth containing 20μg/mL chloramphenicol. These cultures were incubated overnight at 37°C in a shaking incubator before their plasmids were extracted using a Spin Mini Prep kit. Two separate restriction digests were carried out on each of the transformants to check that the ligation was successful. Both the *Not*I / *Xho*I and the *Hind*III digests, for each of the four transformants, were run on an agarose gel (Figure 4.13). The *Not*I / *Xho*I digest cuts the hSTf back out of pALTER-max and *Hind*III digestion should generate three bands (3918bp, 2187bp and 1502bp). All four transformants were shown to contain hSTf-pALTER-max.

One of these transformants was selected for use in the creation of the hSTf-modpET21d expression vector. However, before being used, the hSTf region of this transformant was DNA sequenced (Cytomyx, Cambridge) to check for gene integrity. The sequence was also used to ensure that no mutations were present in the coding frame which would cause an amino acid substitution. Only three silent mutations were found.

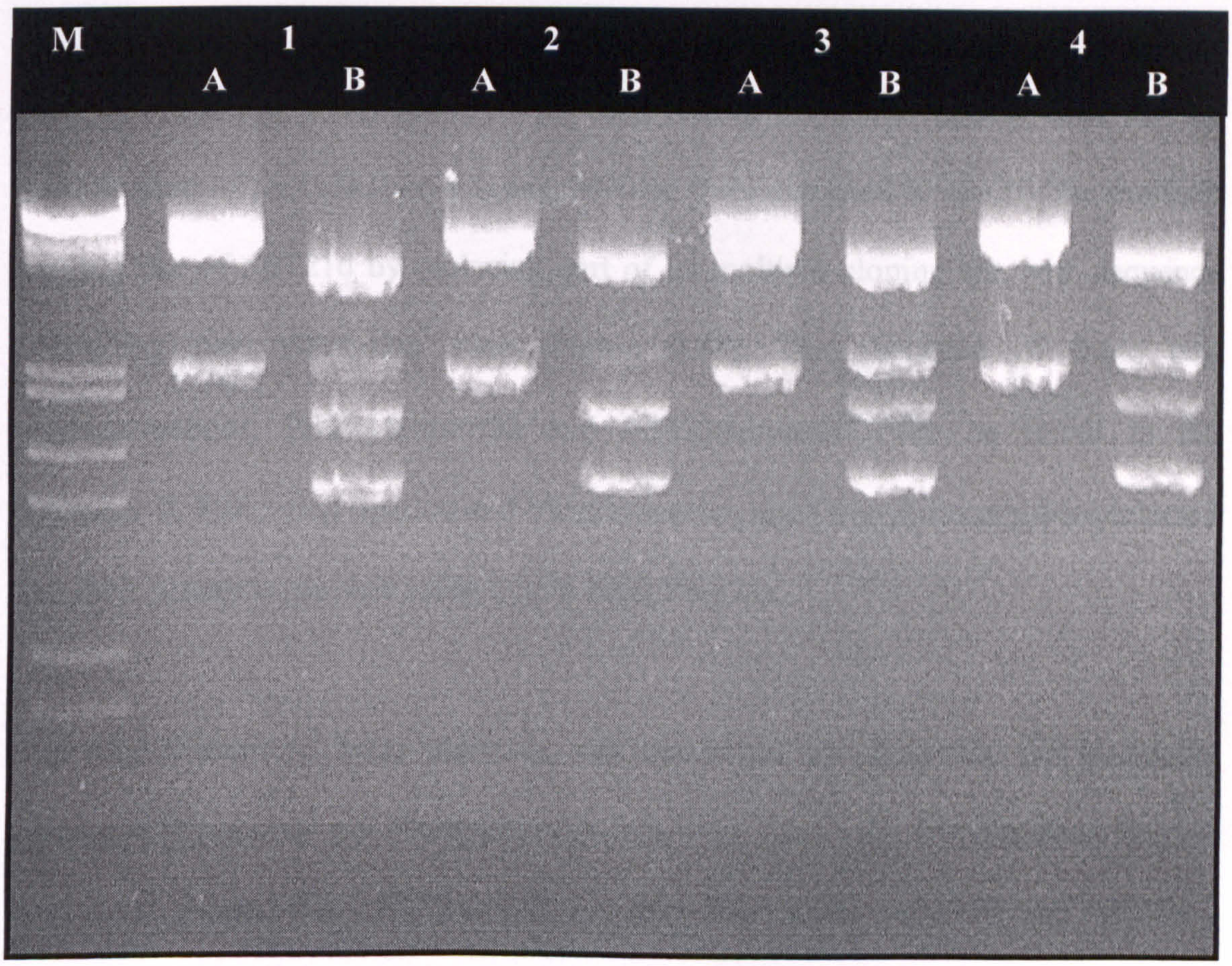


Figure 4.13. Agarose gel electrophoresis of *Not*I / *Xho*I and *Hind*III hLf pALTER-max digests. M: *Eco*R1 / *Hind*III lambda marker, 1-4: hLf pALTER-max transformants, A: *Not*I / *Xho*I digest and B: *Hind*III digest.

4.1.3.3. Creation of hSTf-modpET21d

Rather than PCR cloning hSTf out of pALTER-max and ligating it into pET21d, a three step cloning strategy was used. There were two reasons for doing this: firstly PCR of any large DNA fragment is prone to introduce mutations and secondly, no suitable restriction sites were available for ligation.

Step one was the modification of pET21d. Using the same method used for the modification of pET12a, two oligonucleotides were designed to modify the multiple cloning site of pET21d by replacing part of its multiple cloning site with sequence containing new restriction sites (Figure 4.14). The sequence of these two oligonucleotides, PET21d-insert53 and PET21d-insert35, can be found in the appendix II. Following ligation and transformation, three transformants were selected and their plasmids purified using a Spin Mini Prep kit.

The modification of these plasmids was confirmed by observing the fragments generated by agarose gel electrophoresis following *Sac*I / *Eco*RV digestion (Figure 4.15). An unmodified pET21d will produce two bands (1321bp and 4119bp), whereas the modified pET21d will only produce one band (5440bp). All three transformants contained the insert, but only one was used in further cloning steps.

The second step was to clone into the modified pET21d (modpET21d) the back $^{7}/_{8}$ of hSTf. The modpET21d plasmid and hSTf pALTER-max were digested with *Bam*HI and *Not*I. *Bam*HI cuts hSTf 266bp in from the start whilst *Not*I cuts hSTf straight after its stop codon. The modpET21d digest was cleaned up using a Mini Elute Reaction Cleanup kit, whilst the fragment of hSTf released from hSTf pALTERmax

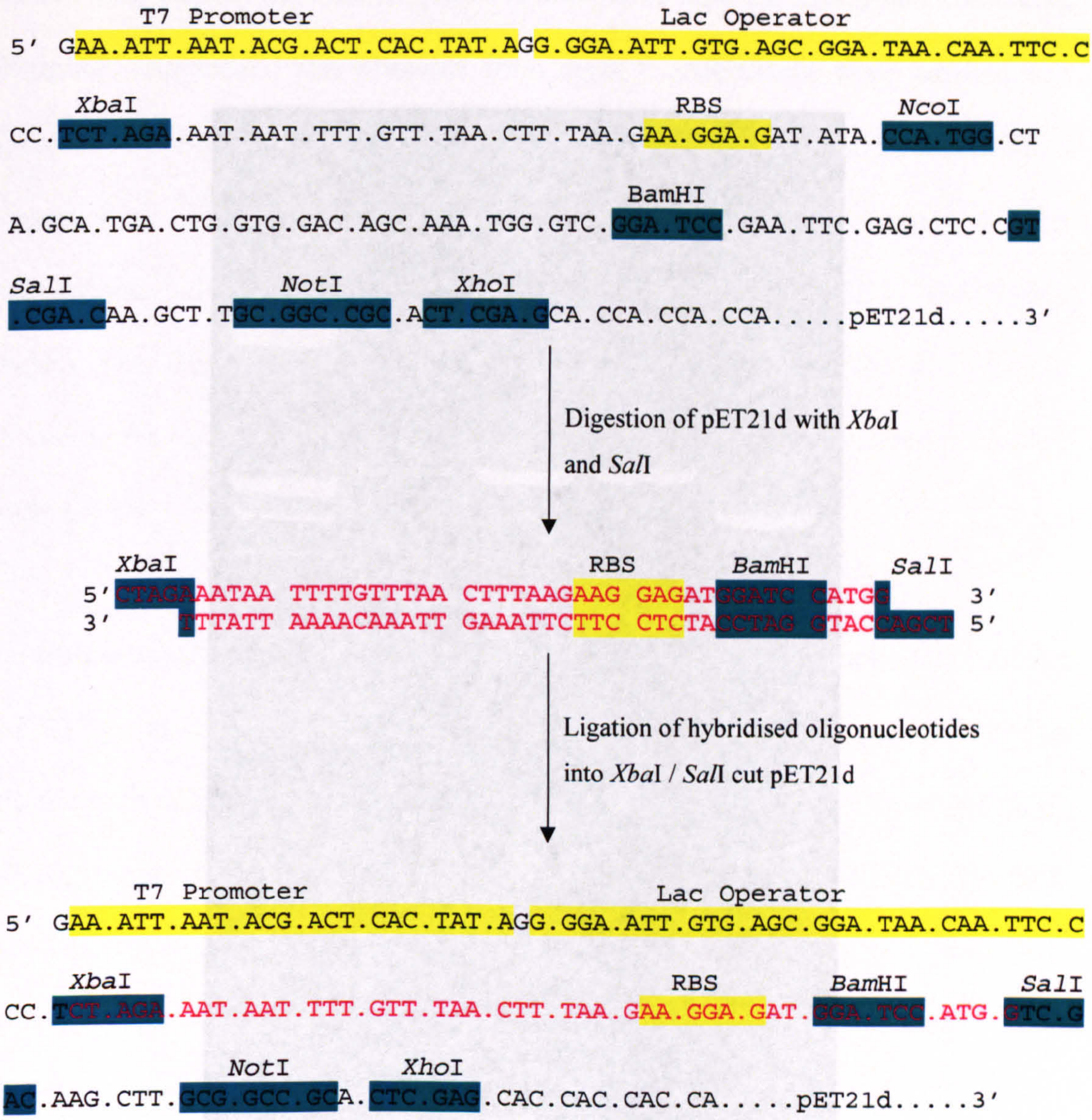


Figure 4.14. Modification of pET21d by replacing part of its multiple cloning site with a 49 base long double-stranded oligonucleotide. Black sequence: unmodified pET21d; red sequence: oligonucleotides; yellow highlighted sequence: regions important for protein expression; and turquoise highlighted sequence: restriction sites.

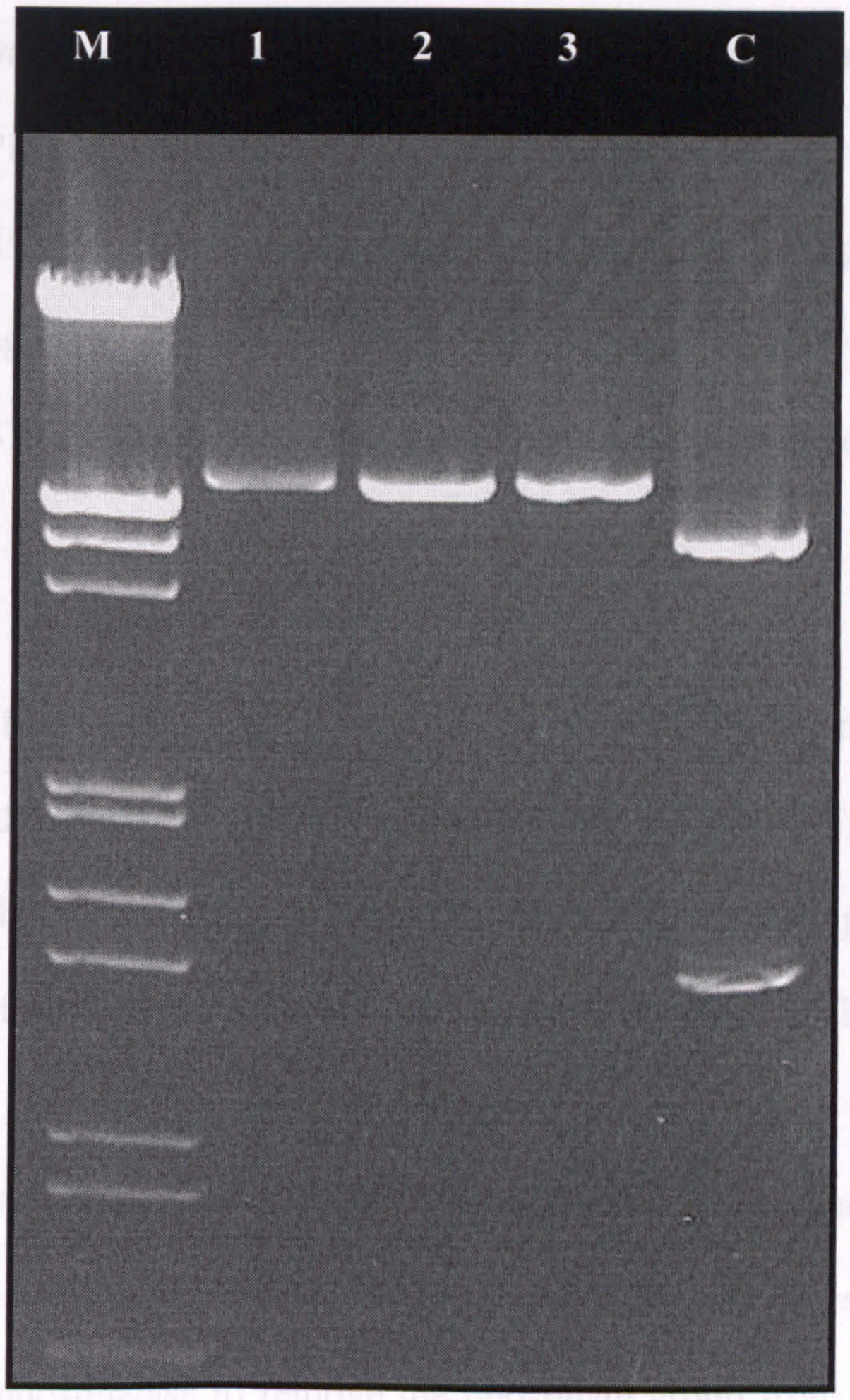


Figure 4.15. Agarose Gel Electrophoresis of three modified pET21d transformants digested with *Sac*I / *Eco*RV. M: *Eco*R1 / *Hind*III lambda marker, 1-3: Transformants and C: pET21d digested with *Sac*I / *Eco*RV

was purified using a Mini Elute Gel Extraction kit. Both were then ligated together before being transformed into JM109 cells and plated onto LB agar plates containing 100μg/mL ampicillin. The plasmids from eight transformants were purified and digested with *Pst*I to confirm the presence of the hSTf fragment in modpET21d (Figure 4.16). Plasmids containing the hSTf fragment will generate a 5211bp fragment and a 1933bp fragment when cut with *Pst*I, whilst modpET21d will only produce one fragment. All of the eight transformants carried the modpET21d containing the hSTf fragment (BhSTfmodpET21d). Only one was selected for further cloning steps.

The final step in the cloning of hSTf into pET21d was the PCR amplification of the first 271bp of hSTf and its ligation into BhSTfmodpET21d (Figure 4.17). The amplification of this region using hSTfp21dFwd and hSTfp21dBck primers generates a 291bp fragment flanked by a *Bam*HI site at the start and a *Bgl*II site at the end. Following PCR amplification of this fragment (section 2.2.5) the product was purified by agarose gel electrophoresis and gel extraction before being digested with *Bam*HI and *Bgl*II. At the same time, BhSTfmodpET21d was digested with *Bam*HI and treated with phosphatase to remove the phosphate group from the 5' end. Both digests were then purified using a Mini Elute Reaction Cleanup kit before being ligated together. The reason for creating a *Bgl*II site is that it can be directly ligated to ends generated using *Bam*HI without affecting the coding frame, but following ligation, the site cannot be recut. Following transformation of this ligation into JM109 cells and overnight growth on LB agar plates containing 100μg/mL ampicillin, 12 transformants were selected and their plasmids purified. All 12 transformants were digested with *Nde*I and then run on an agarose gel (Figure

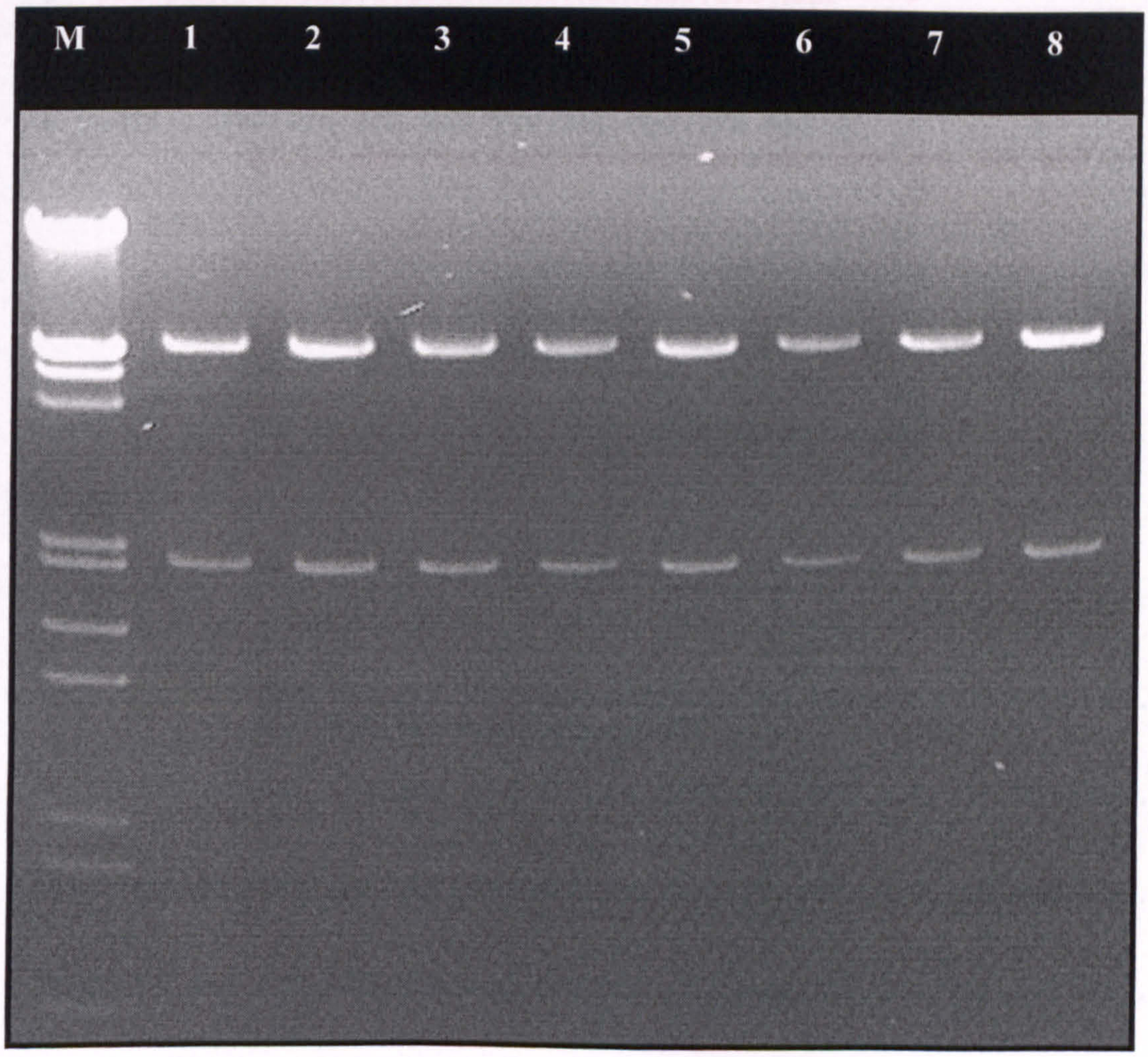


Figure 4.16. Agarose Gel Electrophoresis of eight modpET21d transformants containing the back $^{7}/_{8}$ of the hSTf cDNA following digestion with *Pst*I. M: *Eco*R1 / *Hind*III lambda marker and 1-8: transformants.

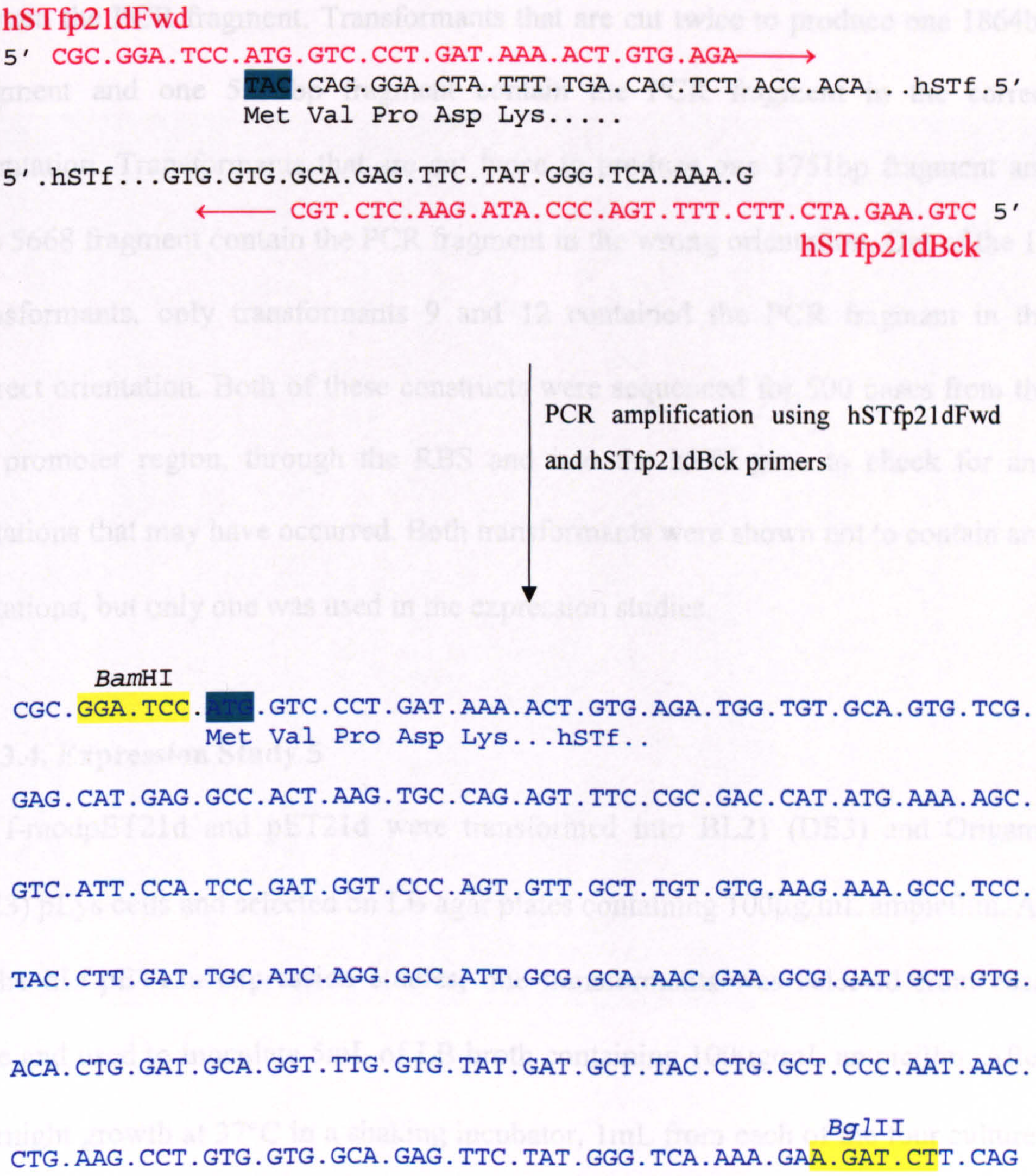


Figure 4.17. PCR amplification of the first 271bp of hSTf and the creation of *Bam*HI and *Bgl*II restriction sites using hSTfp21dFwd and hSTfp21dBck primers. Black sequence: hSTf pALTER-max; red sequence: oligonucleotides; blue sequence: PCR product; yellow highlighted sequence: generated restriction sites and turquoise highlighted sequence: hSTf start codon.

4.18) to determine the orientation of the PCR insert. Following *Nde*I digestion, three different plasmids can be visualised. Transformants that are only cut once do not contain the PCR fragment. Transformants that are cut twice to produce one 1864bp fragment and one 5555bp fragment contain the PCR fragment in the correct orientation. Transformants that are cut twice to produce one 1751bp fragment and one 5668 fragment contain the PCR fragment in the wrong orientation. Out of the 12 transformants, only transformants 9 and 12 contained the PCR fragment in the correct orientation. Both of these constructs were sequenced for 500 bases from the T7 promoter region, through the RBS and into the hSTf gene to check for any mutations that may have occurred. Both transformants were shown not to contain any mutations, but only one was used in the expression studies.

4.1.3.4. Expression Study 5

hSTf-modpET21d and pET21d were transformed into BL21 (DE3) and Origami (DE3) pLys cells and selected on LB agar plates containing 100μg/mL ampicillin. As in the hLf-pET12a expression studies, one transformants was selected from each plate and used to inoculate 5mL of LB broth containing 100μg/mL ampicillin. After overnight growth at 37°C in a shaking incubator, 1mL from each of the four cultures was tested for the presence of its expression vector using a Spin Mini Prep kit, leaving the remaining culture for expression studies. Once all had been confirmed to contain an expression vector, three different inductions were set up for each of the four cultures (Table 4.2 numbers 32-43). After induction, cells were harvested, sonicated and their protein expression profile visualised on an SDS-PAGE (Figure 4.19 numbers 32-43) using the same method as used for all the previous expression studies.

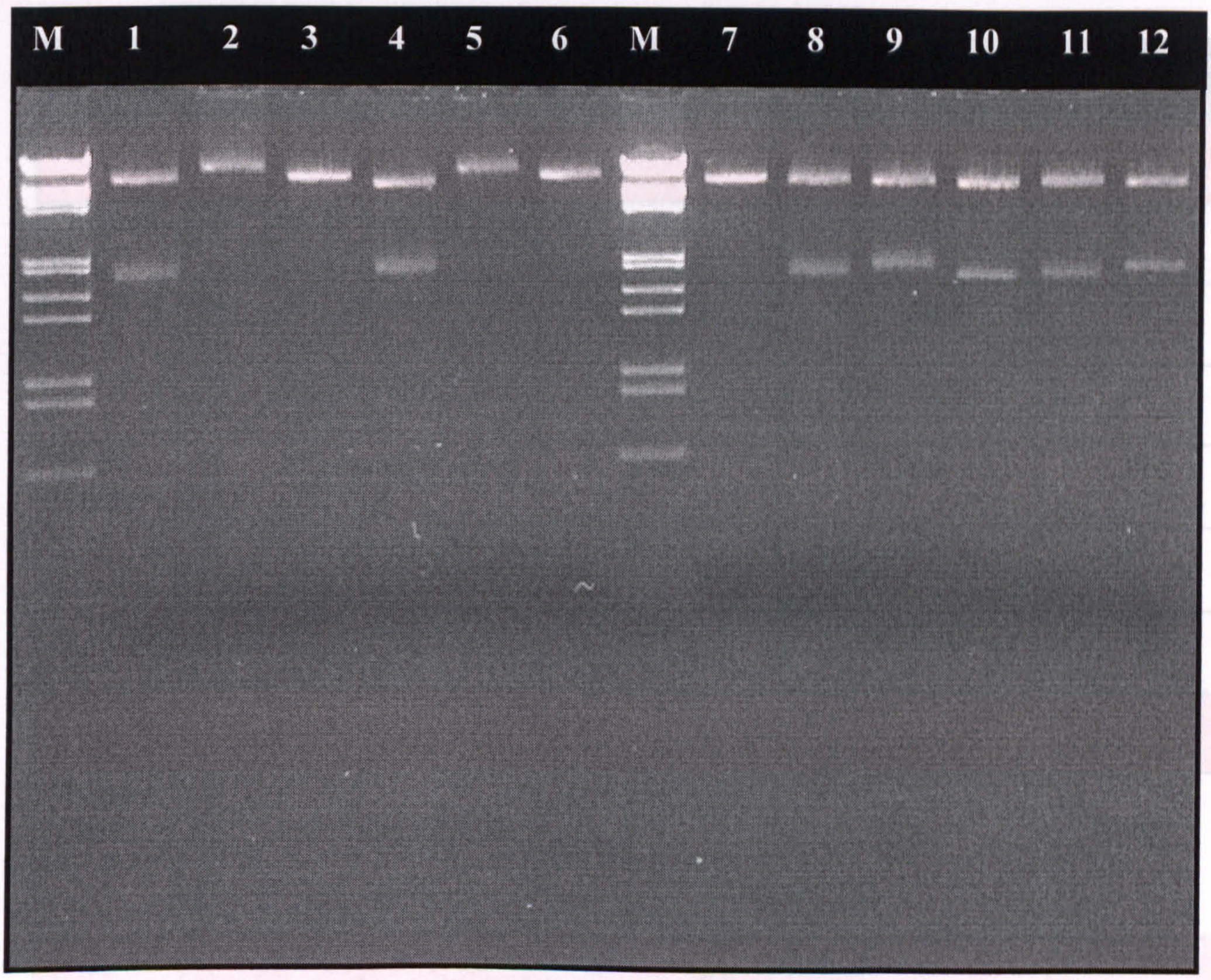


Figure 4.18. Agarose gel electrophoresis of *Nde*I digested hSTfmodpET21d transformants. M: *Eco*R1 / *Hind*III lambda marker and 1-12: *Nde*I digested transformants.

E. coli Strain	No.	Expression Construct	Pre-Induction Growth	Post-Induction Growth	Incubation Temp (°C)
Expression Study 5: IPTG Induction					
Origami (DE3)	32	hSTfmodpET21d	1 hrs	4 hrs	37
	33	pET21d			
	34	hSTfmodpET21d	0 hrs	Overnight	37
	35	pET21d			
	36	hSTfmodpET21d	0 hrs	Overnight	25
	37	pET21d			
BL21 (DE3)	38	hSTfmodpET21d	1 hrs	4 hrs	37
	39	pET21d			
	40	hSTfmodpET21d	0 hrs	Overnight	37
	41	pET21d			
	42	hSTfmodpET21d	0 hrs	Overnight	25
	43	pET21d			
Expression Study 6: IPTG Induction					
Origami (DE3)	44	hSTfmodpET21d	3 hrs	2 hrs	37°C
	45		4 hrs		
	46		5 hrs		
	47	pET21d	5 hrs		
Expression Study 7: Bacteriophage λCE6 Induction					
Origami (DE3)	48	hSTfmodpET21d	2 hrs	3 hrs	37
	49	pET21d			
	50	hSTfmodpET21d	2 hrs	5 hrs	
	51	pET21d			

Table 4.2. Induction plan for hSTf expression using pET21d in *E. coli* Origami (DE3) pLys, BL21 (DE3) and Origami cells.

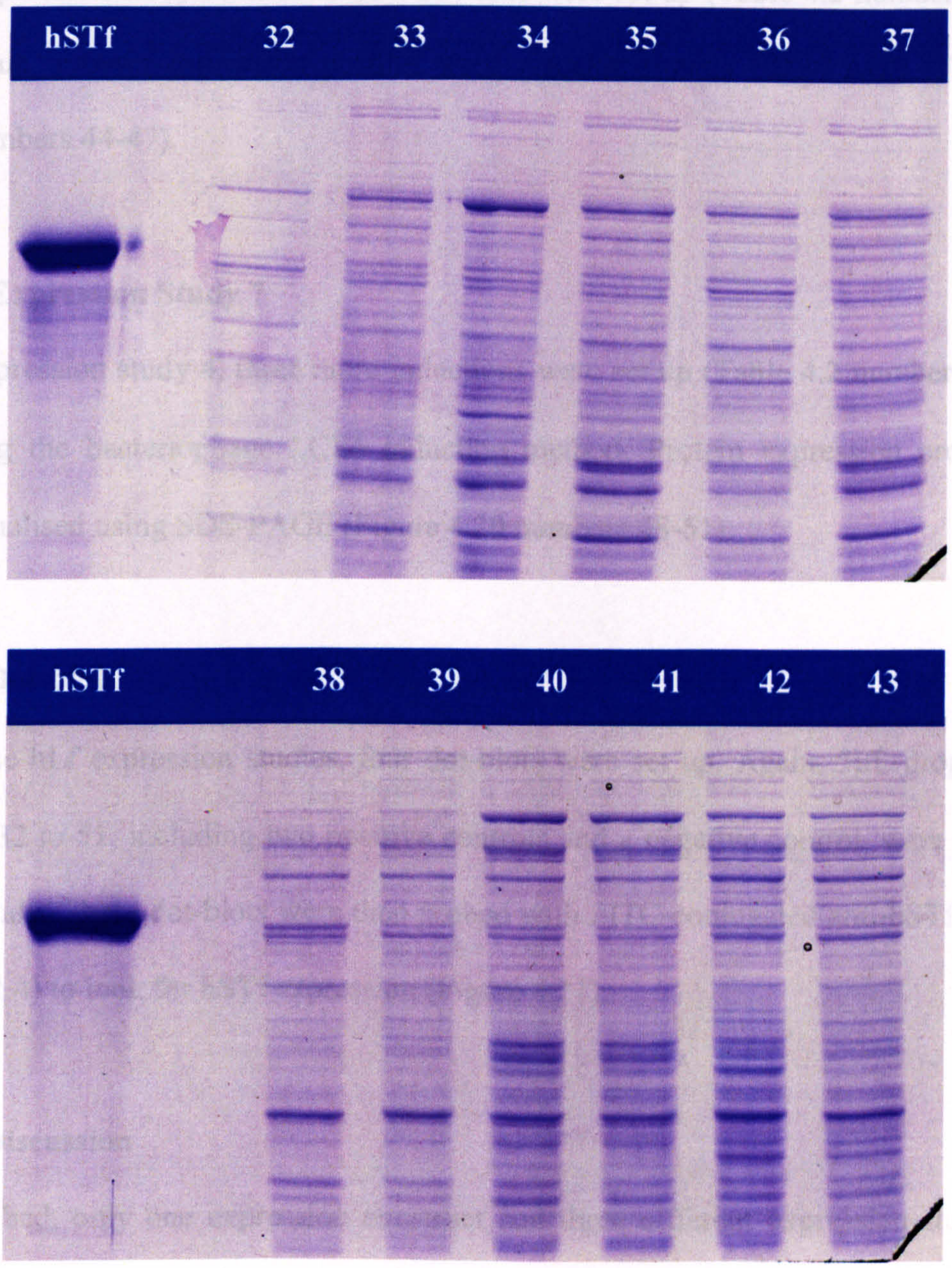


Figure 4.19. SDS-PAGE, under reducing conditions, of IPTG induction experiments, numbers 32-43. Native hSTf was used as a reference band in order to locate any *E. coli* expression of hSTf. Numbers 32-37: Origami (DE3) plys, 38-43: BL2 (DE3) (table 4.2 numbers 32-43). All samples run were extracted from their soluble fractions.

4.1.3.5. Expression Study 6

As in expression study 3, three more inductions were set up (Table 4.2 numbers 44-47). Again, protein expression profiles were visualised using SDS-PAGE (Figure 4.20. numbers 44-47).

4.1.3.6. Expression Study 7

As in expression study 4, three more inductions were set up (Table 4.2 numbers 48-51) using the bacteriophage λCE6 induction method. Protein expression profiles were visualised using SDS-PAGE (Figure 4.20 numbers 48-51).

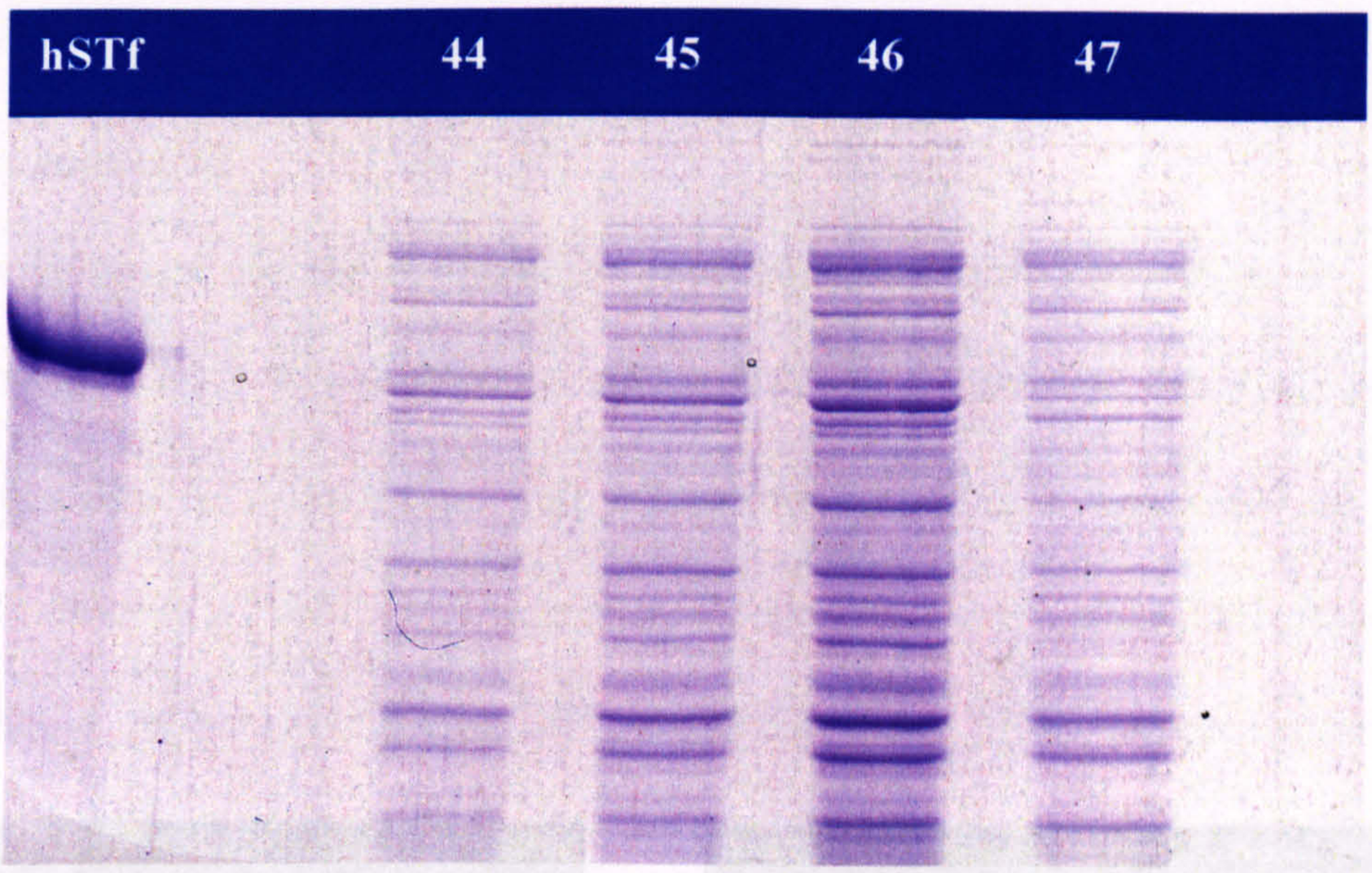
4.1.3.7. Dot-blots

As for the hLf expression studies, four dot-blots were set up. Again, 3μL drops of samples 32 to 51, including two positive controls and a negative control, were used in this study. These dot-blots were then probed with FITC-conjugated anti-hSTf IgG (section 2.4) to look for hSTf expression (Figure 4.21).

4.1.3.8. Discussion

As described, only one expression construct and three different expression studies were explored to look for expression of hSTf in *E.coli*.

Similar expression studies to the ones explored for the expression of hLf in *E. coli* were used for the expression of hSTf in *E. coli*. Expression studies five and six explored IPTG induction, whereas expression study seven used the bacteriophage induction system described before. As for hLf, no expression could be detected in any of these three studies, either using SDS-PAGE or Dot-blot analysis.



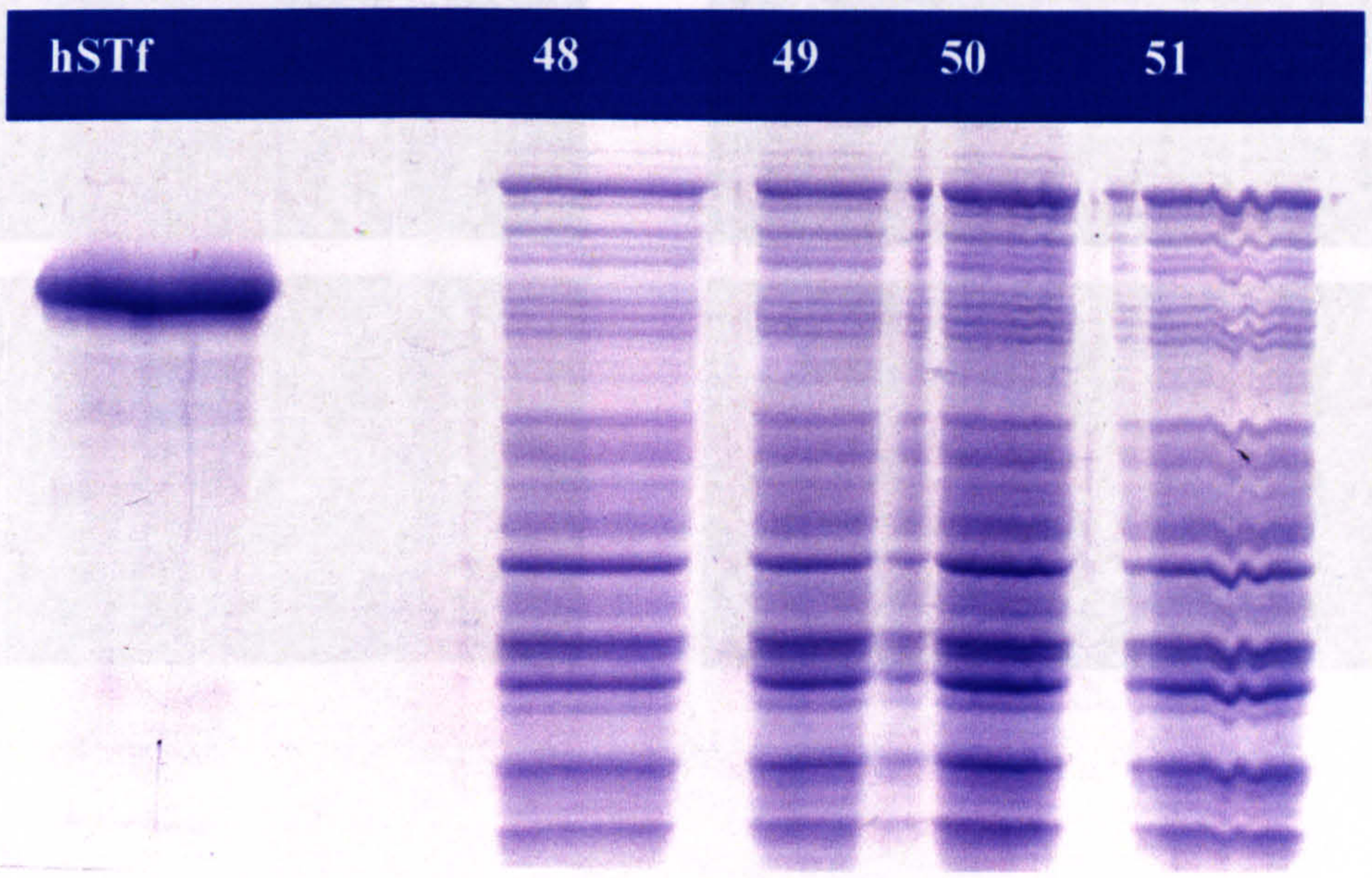


Figure 4.20. Reducing SDS-PAGE of hSTf expression studies 6 and 7, experiment numbers 44-51. hSTf was used as a reference band in order to locate any *E. coli* expression of hSTf. Numbers 44-47: Origami (DE3) plys IPTG induction, 48-51: Origami (DE3) λCE6 induction (table 4.2). All samples run were extracted from their soluble fractions.

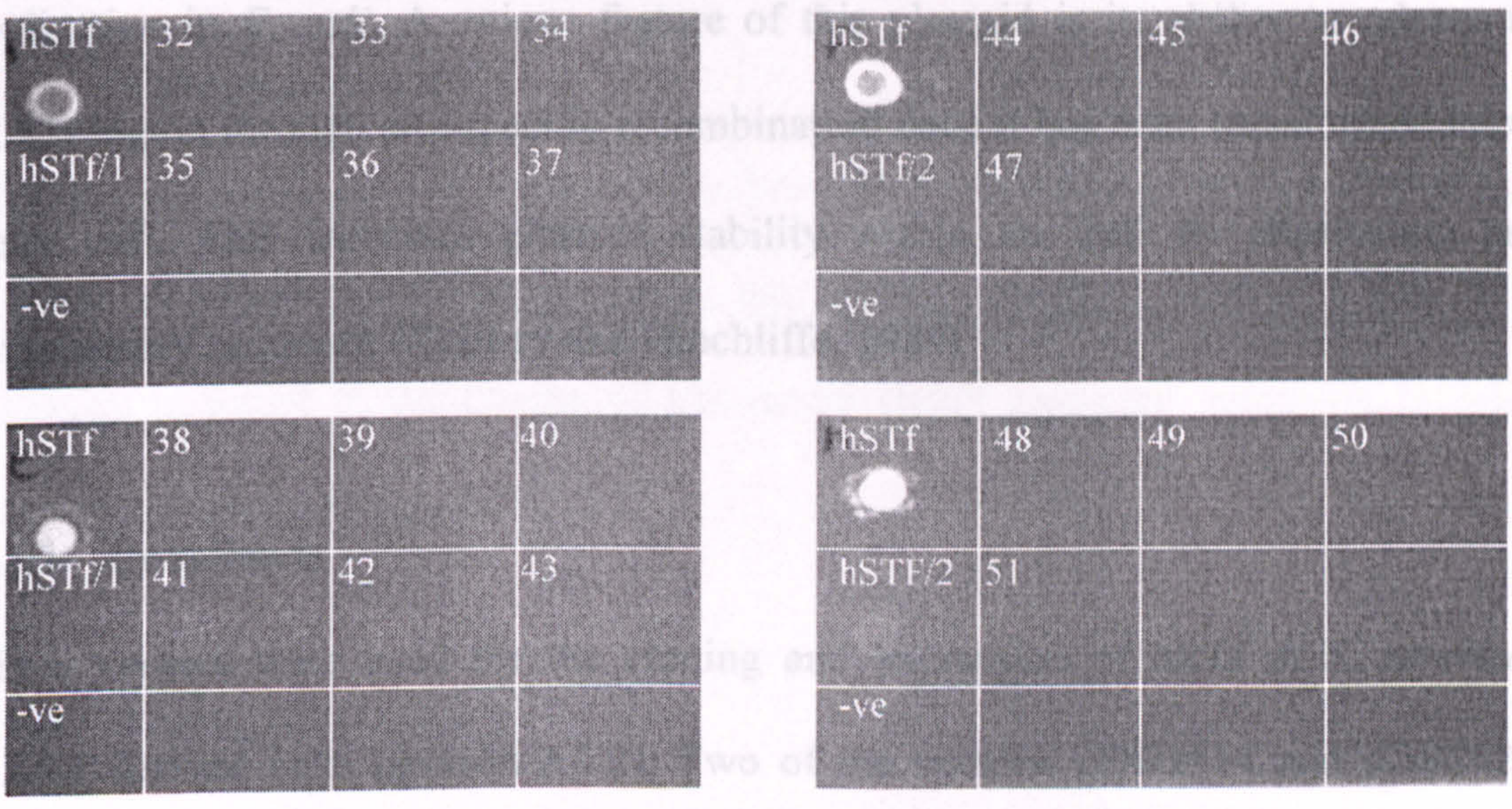


Figure 4.21. Dot Blot analysis of samples 32 to 51 with FITC-conjugated Anti-hSTf IgG to look for hSTf expression. hSTf: 1mg/ml hSTf positive control, hSTf/2: 0.5mg/ml hSTf positive control, -ve: water and 32-51: samples from expression studies 5,6 and 7 (table 4.2).

4.2. Expression of hSTf in *Saccharomyces cerevisiae* and mutagenesis

4.2.1. Introduction

Using an in-house protein expression system designed and constructed by Delta Biotechnology Ltd, it was possible to express hSTf in *S. cerevisiae* at levels greater than 1g per litre in a fermentor. The expression vector used was based on a 2μ plasmid containing a *PRB1* yeast promoter, a fusion leader for protein secretion, the hSTf insert, a gene encoding ampicillin resistance and a pUC9 region for plasmid replication in *E. coli*. A unique feature of this plasmid is its ability to release its pUC9 region through site-specific recombination once it has been transformed into a yeast cell. This increases plasmid stability within the cell by eliminating any unnecessary sequence (Chinery and Hinchliffe, 1989).

4.2.2. Constructs

Three vectors were used for the cloning and expression of hSTf in *S. cerevisae* DXY1 derived cells (section 2.1.2). Two of the vectors, pDB2514 and pDB2529, were used purely for cloning and modification steps, whilst pDB2536 was the actual expression vector (section 2.1.7). All three vectors contained the hSTf cDNA insert (appendix II).

In order to produce a homogenous population of deglycosylated proteins, which would improve the crystallisation conditions (Pace, 1990; Mason *et al.* 2001), the double mutant (N413Q, N611Q) was generated using site-directed mutagenesis kindly donated by Christopher Finnis (Delta Biotechnology). Insertion of both mutations is unlikely to compromise the structural integrity of the protein or its

function, since previous studies have shown that the glycan chains have no functional role in transferrin (Mason *et al.* 1993). The reason for substituting an asparagine with a glutamine is due to their close structural similarity, as shown by Dayhoff *et al*, 1978, and as discussed in section 3.2.4, the fact that these mutations are predicted not to affect the structure or function of the protein. Mason *et al*, 1998 were also able to express non-glycosylated human transferrin in BHK cells following mutagenesis of the hSTf cDNA. However, instead of mutating N413 and N611 to two glutamines, they mutated these two residues to aspartates.

4.2.3. Mutagenesis of hSTf

The reason behind developing an hSTf expression system was to be able to express a number of specific mutants, primarily the G394R mutation (section 1.3.9). This mutation is found in the C-lobe of transferrin (hSTf sequence, appendix I), lying close to the iron binding-site.

4.2.3.1. Mutagenesis Design

Mutagenesis of the hSTf construct was achieved using a QuikChange site-directed mutagenesis kit from Stratagene. To create a substitution from glycine to an arginine at position 394 two oligonucleotides (CF125 and CF124) were designed (appendix II). CF125 was designed to bind to a 35 bases region around codon 394 on the 5' to 3' strand, whilst CF124 was designed to bind the same region but on the 3' to 5' strand (Figure 4.22). Both oligonucleotides complement a region from codon 388 to 399 of hSTf, except codon 394 where two bases have been changed so that an arginine is encoded rather than a glycine. The reason for mutating codon 394 from GGG to AGA, and not to any other codon encoding arginine, is because AGA is the

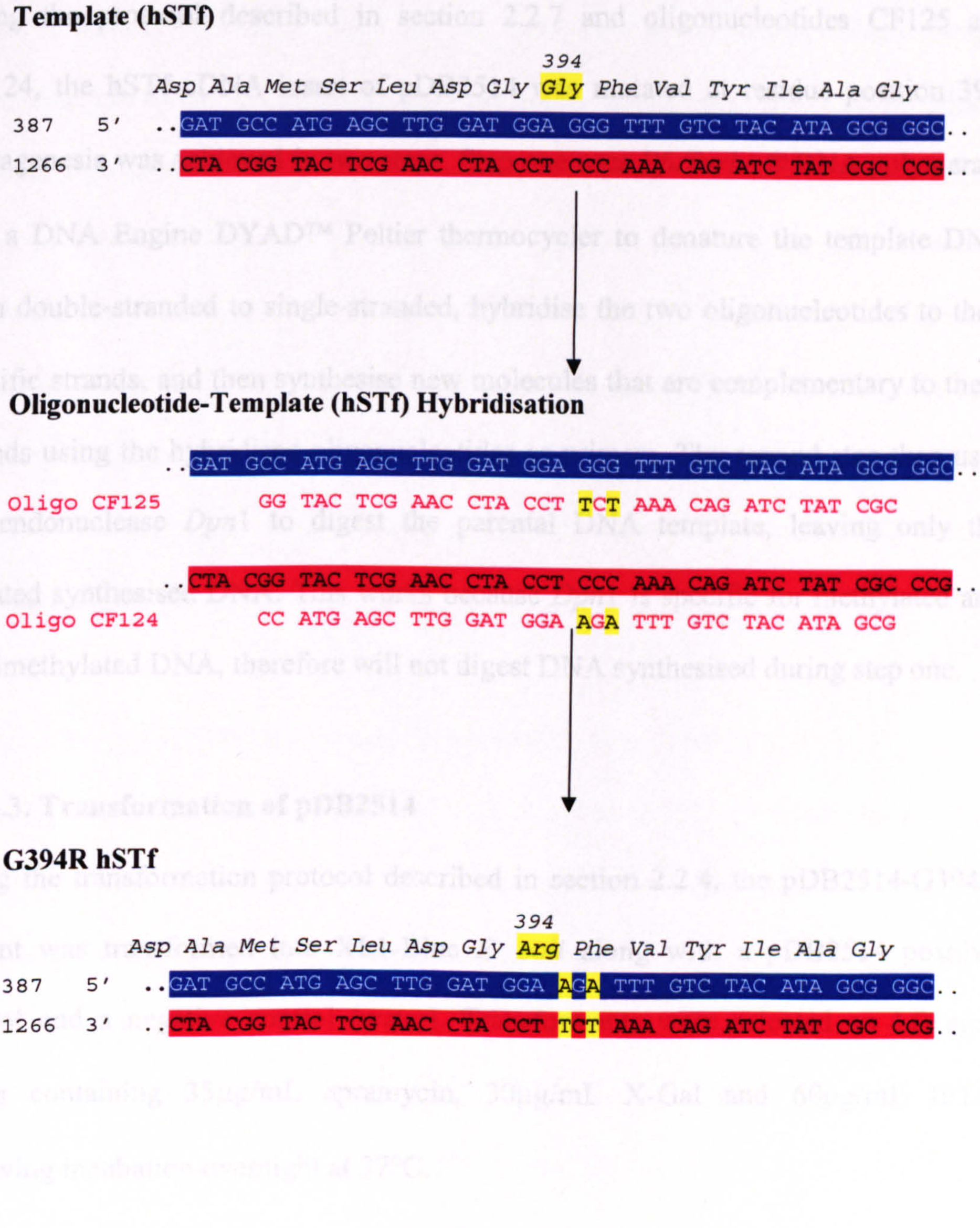


Figure 4.22. Creation of the G394R mutation in hSTf using a Stratagene QuikChange site-directed mutagenesis kit. Mutated residues are highlighted in yellow.

most commonly used codon by *S.cerevisae* to encode arginine (ftp://genome-ftp.stanford.edu/pub/codon/ysc.orf.cod).

4.2.3.2. Mutagenesis

Using the protocol described in section 2.2.7 and oligonucleotides CF125 and CF124, the hSTf cDNA insert of pDB2514 was mutated at residue position 394. Mutagenesis was achieved in two steps. Step one uses *PfuTurbo®* DNA polymerase and a DNA Engine DYAD™ Peltier thermocycler to denature the template DNA from double-stranded to single-stranded, hybridise the two oligonucleotides to their specific strands, and then synthesise new molecules that are complementary to these strands using the hybridised oligonucleotides as primers. The second step then uses the endonuclease *Dpn*1 to digest the parental DNA template, leaving only the mutated synthesised DNA. This works because *Dpn*1 is specific for methylated and hemimethylated DNA, therefore will not digest DNA synthesised during step one.

4.2.3.3. Transformation of pDB2514

Using the transformation protocol described in section 2.2.4, the pDB2514-G394R mutant was transformed into XL1-Blue *E. coli* along with a pDB2514 positive control and a negative control (water). Transformants were selected on LB agar plates containing 35μg/mL apramycin, 30μg/mL X-Gal and 60μg/mL IPTG following incubation overnight at 37°C.

4.2.3.4. Mutagenesis Check

To check that the desired mutations have been made, six transformant colonies were picked and individually grown at 37°C overnight in 100mL of LB broth containing 35μg/mL apramycin. The pDB2514-G394R constructs were then purified from each clone using a Qiagen HiSpeed Plasmid Midi Kit (section 2.1.8). Plasmids were eluted in 350μL of MilliQ water and the DNA concentration determined using a Shimadzu UV-1201 UV-VIS Spectrophotometer measuring the OD at 260nm (section 2.2.6). A 728bp sequence between restriction sites *Asp*718I and *Bsp*E1 found in the hSTf insert of each pDB2514-G394R clone was sequenced with a Perkin Elmer ABI Prism 310 Genetic Analyser (section 2.2.8) using six specific oligonucleotides to confirm the two base mutations in codon 394 and to make sure no other mutations have occurred. Out of the six transformants, five of them contained the desired mutations, but only one transformant was selected for further cloning experiments.

4.2.3.5. Cloning of G394R fragment from pDB2514 into pDB2529

The selected transformant was digested with *Asp*718I and *Bsp*E1 to release a 728bp fragment. At the same time, pDB2529 was digested with the same two enzymes. Following electrophoresis of the two digests on a 1% (w/v) agarose gel (Figure 4.23), a BIO101 Gene Clean III gel extraction kit (section 2-1-8) was used to extract the 728bp fragment from pDB2514 and the 5757bp fragment from pDB2529. Once purified, these two fragments were ligated together (section 2.2.3) before being transformed into DH5α cells and plated onto LB agar plates containing 50μg/mL ampicillin. Following overnight incubation at 37°C, two transformants were selected.

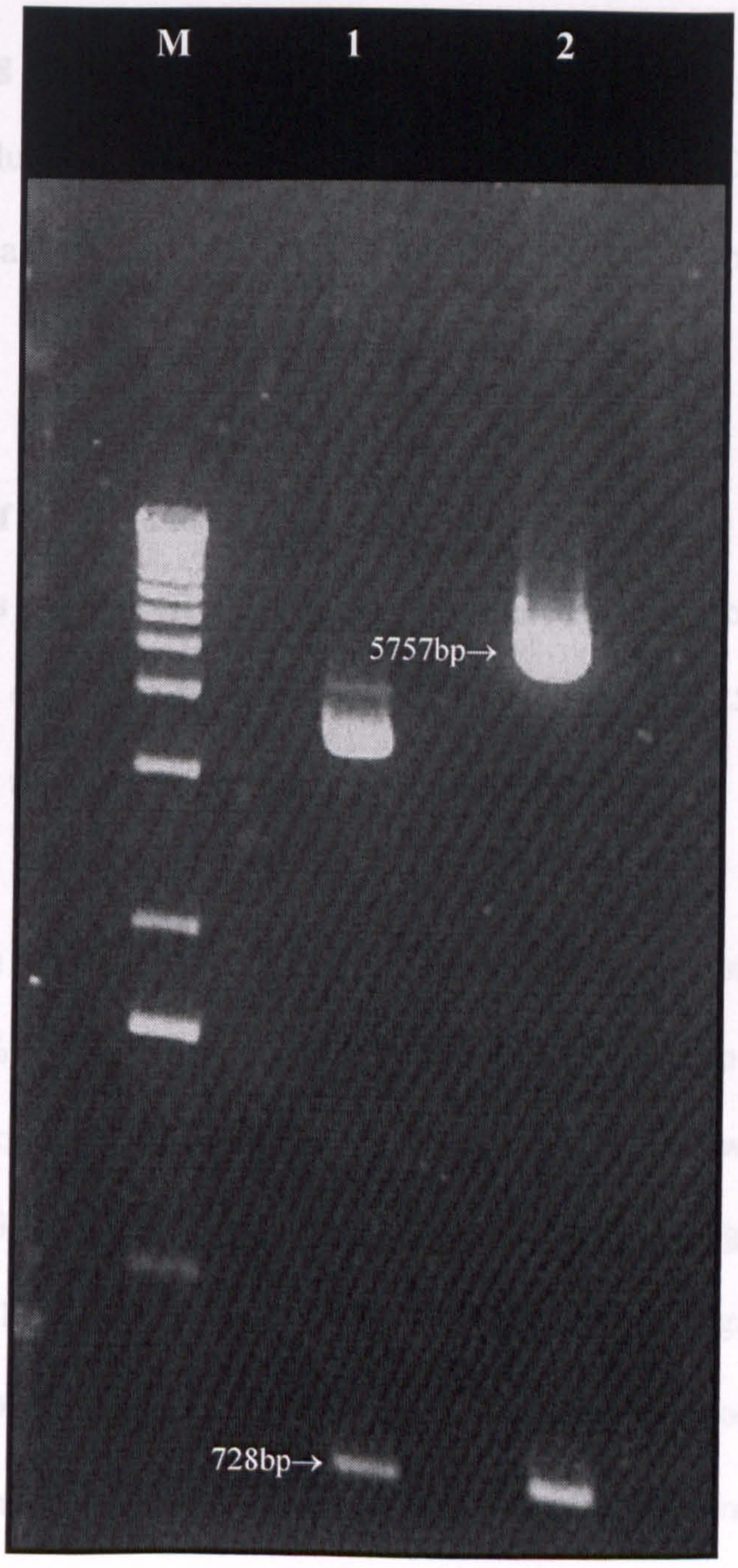


Figure 4.23. DNA gel electrophoresis of pDB2514 and pDB2529 following digestion with Asp718I and BspE1. M: 1Kb Marker, 1: pDB2514 and 2: pDB2529.

The modified plasmids from both pDB2529 transformants were purified using a Qiagen HiSpeed Plasmid Midi Kit following their inoculation into 100mL LB cultures containing 50μg/mL ampicillin and overnight incubation at 37°C. As before, Plasmids were eluted in 350μL of MilliQ water and the DNA concentration determined using a Shimadzu UV-1201 UV-VIS Spectrophotometer measuring the OD at 260nm.

4.2.3.6. Cloning of G394R Insert from pDB2529 into pSAC35

In order to express the G394R hSTf gene it was necessary to clone it together with all other features of the NotI expression cassette from pDB2529, into pSAC35 (pDB2536 without the NotI hSTf expression cassette).

The plasmids from both pDB2529 transformants were digested using *Not*I and *Sca*I and run on a 1% (w/v) agarose gel (Figure 4.24). The two identical 3274bp fragments, corresponding to the *Not*I hSTf expression cassette, were then extracted using a BIO101 Gene Clean III gel extraction kit before being ligated into *Not*I digested and dephosphorylated pSAC35 plasmids. Both ligations were then transformed into DH5α cells and plated onto LB agar plates containing 50μg/mL ampicillin, 30μg/mL X-Gal and 60μg/mL IPTG. Following incubation of these transformants overnight at 37°C, 96 colonies were picked and used to inoculate a 96 well plate containing 1.3mL of LB broth and 50μg/mL ampicillin per well. This block was incubated overnight on a shaking platform at 37°C. A plasmid prep of each transformant was then carried out using a Qiaprep 96 Turbo kit (section 2-1-8).

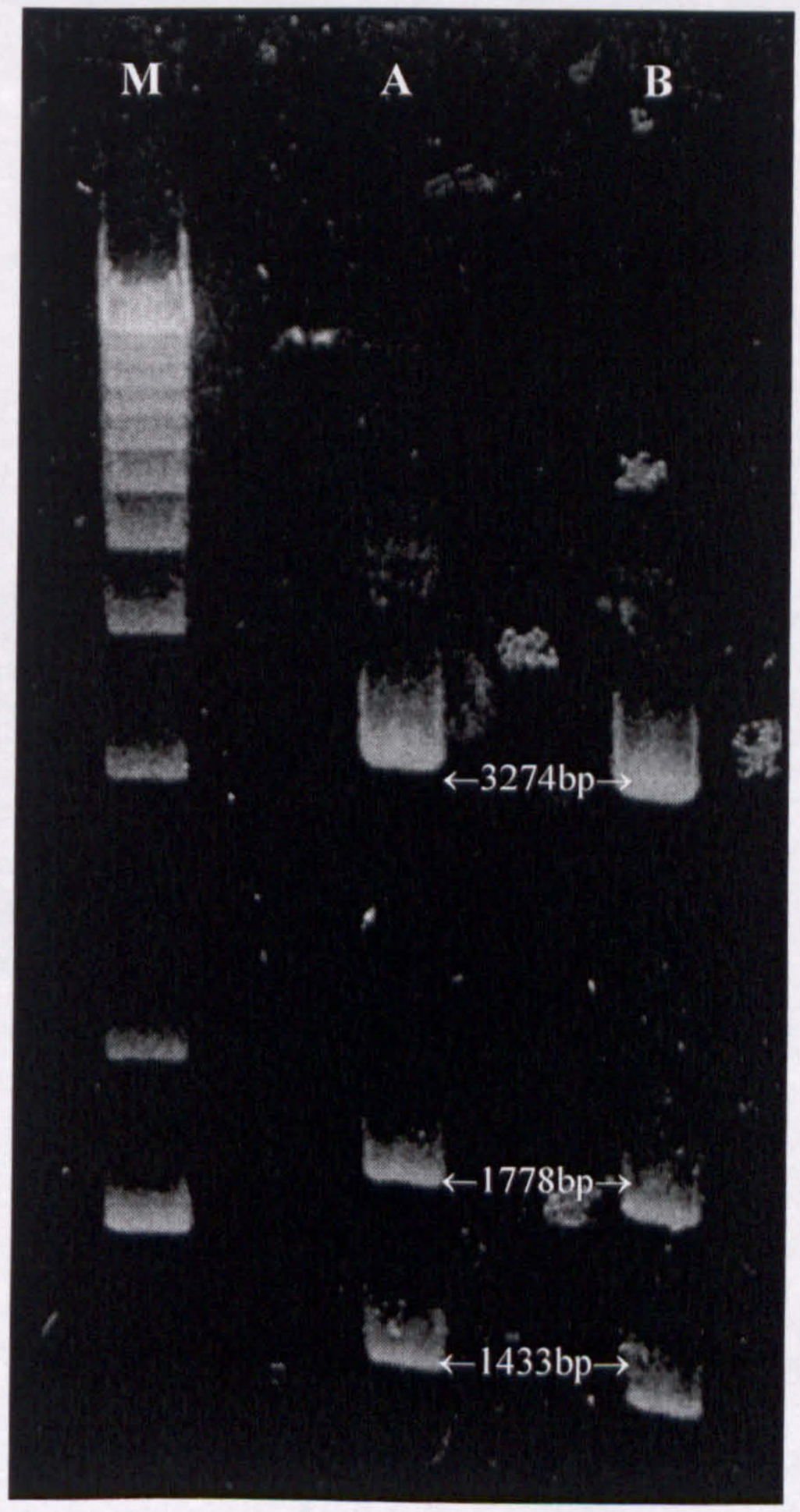


Figure 4.24. DNA Gel Electrophoresis of G394R pDB2529, following *Not*I / *Sca*I digestion. M: 1Kb Marker, A: pDB2529 transformant 1 and B: pDB2529 transformant 2.

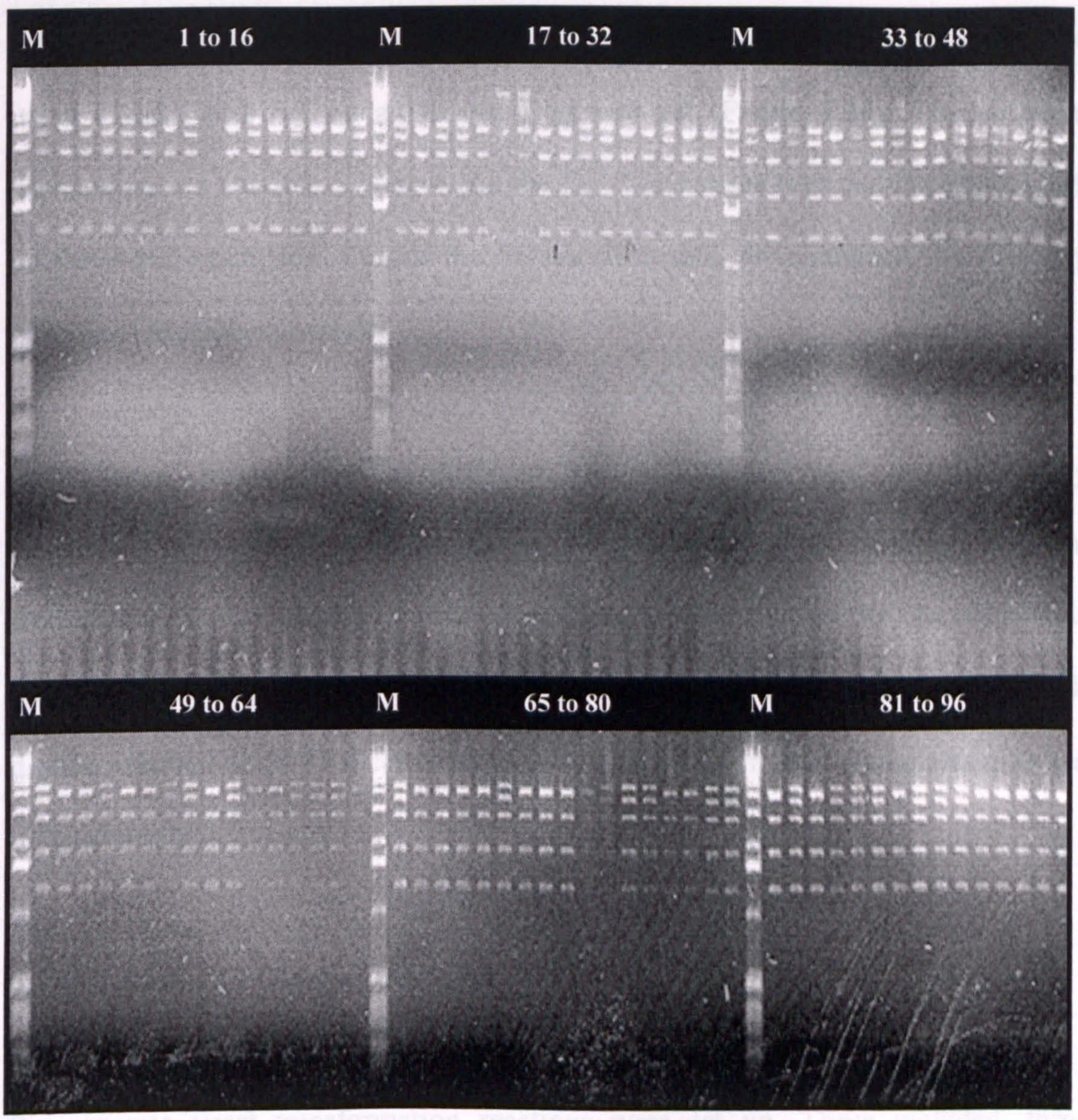


Figure 4.25. DNA gel electrophoresis of G394R pDB2536, following *Hind*III digestion. M: 1Kb Marker. The 96 plasmid prep digests were run in groups of 16 with a marker lane in between.

To check whether the *Not*I hSTf expression cassette has been ligated into pSAC35 in the correct orientation, 5μL samples of each of the 96 plasmid preps were digested using *Hind*III and run on a 1% (w/v) agarose gel (Figure 4.25). Two different orientations are possible (Figure 4.26): the *LEU2* orientated (optimal for expression), or the AMP orientated. Out of the 96 transformants, 44 contained the correctly orientated expression cassette.

From these results, three correctly orientated constructs and two incorrectly orientated constructs were picked. Carrying out an RFLP study using five different endonucleases, the integrity of these five plasmids were checked by electrophoresis on a 1% (w/v) agarose gel (Figure 4.27). The theoretical fragment sizes for all digests matched up to what can be seen on the gel, showing that all the plasmids had maintained their integrity.

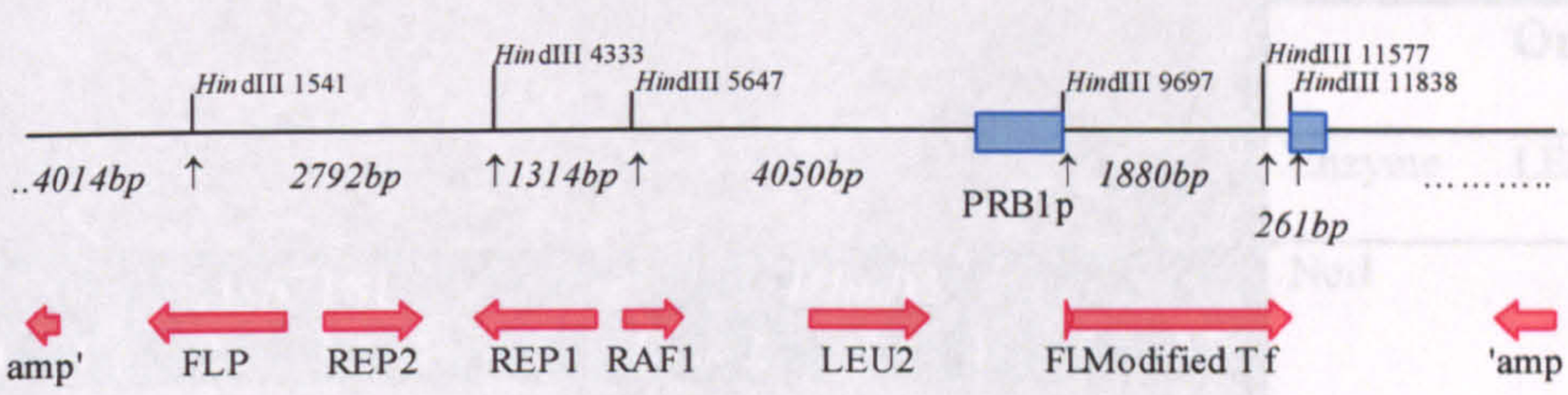
4.2.4. Transformation and Stock Set up

4.2.4.1. Yeast Transformation

Six separate yeast transformation reactions were set up. These included, three correctly orientated pDB2536-hSTf plasmids (section 4.2.3), two positive controls (pSAC35 and a Sigma control vector pR536) and a negative control (no DNA).

Following transformation (section 2.2.4) five colonies from each of the pDB2536-hSTf plates and five colonies from the pSAC35 plate were used to inoculate separate flasks containing 10mL BMM-D broths. Using the same twenty colonies, each colony was streaked in a patch on BMM-D agar plates. The flasks were shaken incubated at 30°C for 4 days and the plates were incubated at 30°C for 2 days.

Product 1: LEU2 Orientated



Product 2: Amp Orientated

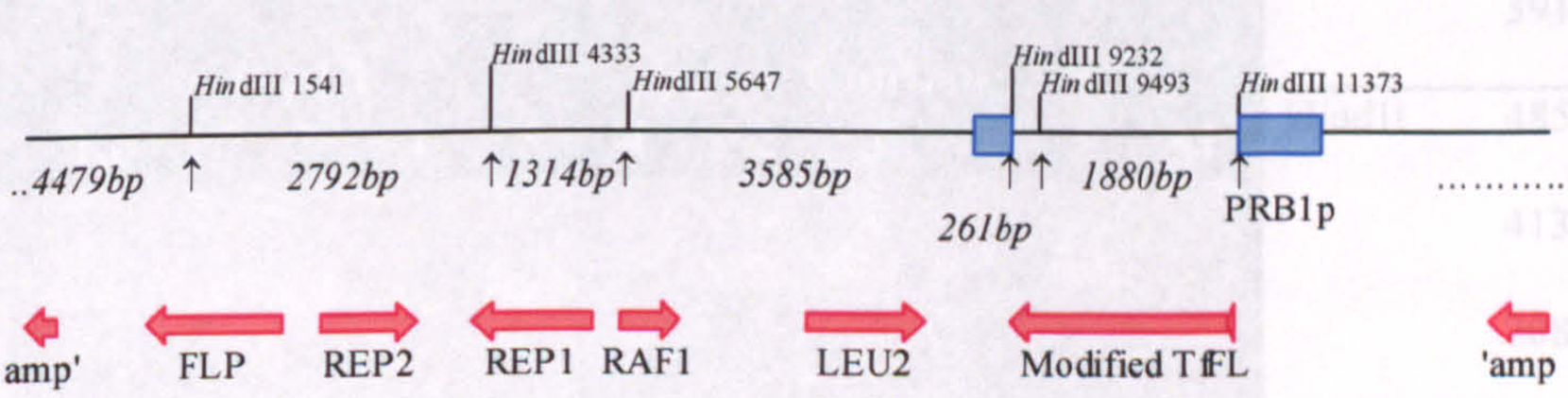
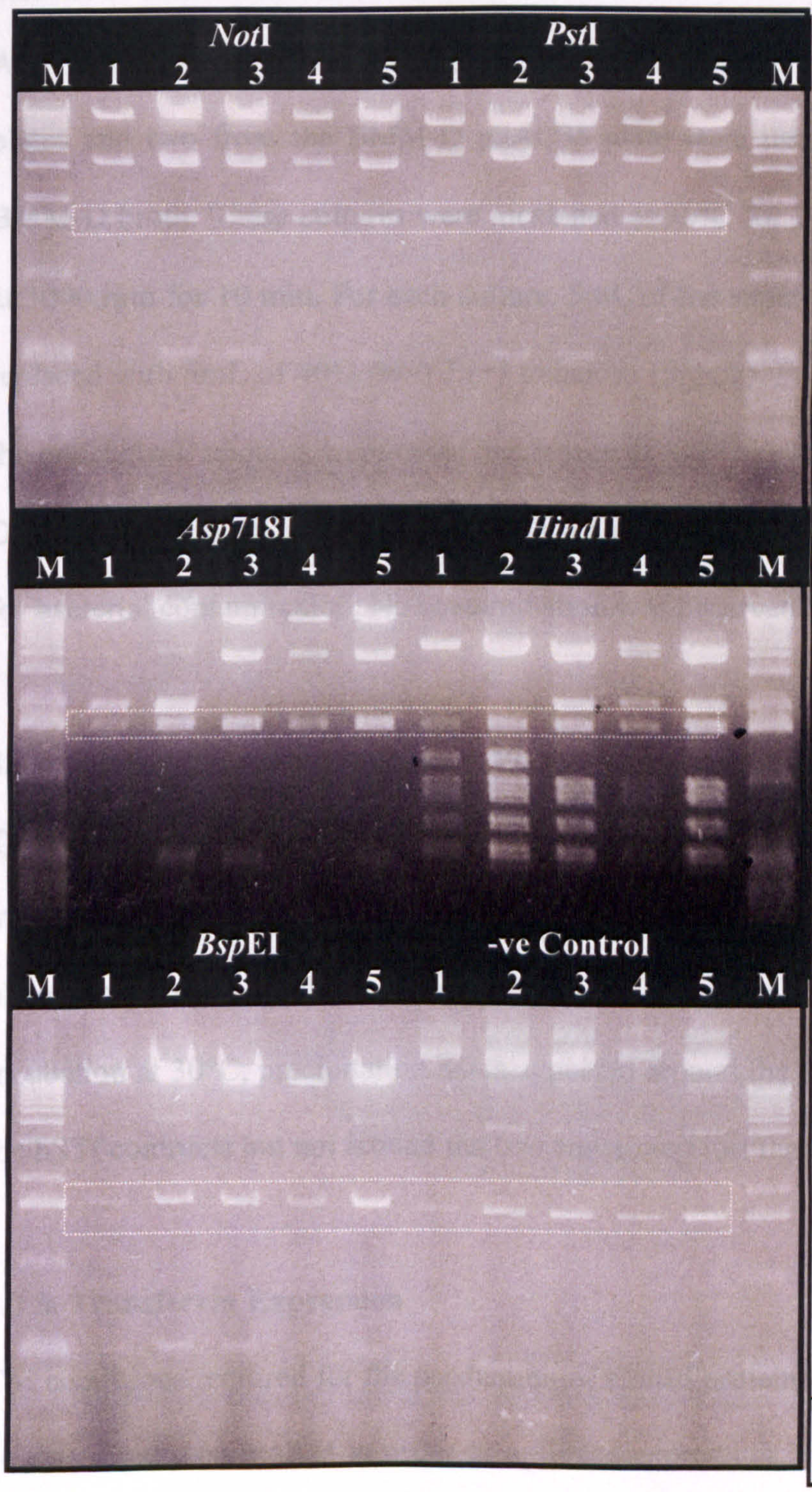


Figure 4.26. Maps of the two possible ligation products following cloning of the *Not*I hSTf expression cassette into pSAC35 and the theoretical fragment sizes produced following *Hind*III digestion. Because the two largest fragments from product one are of a similar size, they will not resolve on a gel and therefore it will appear that there is only one fragment.



Fragment sizes produced following digestion of five pDB2536 transformants (bp)		
	Orientation	
Enzyme	LEU2	Amp
NotI	11037 3274	
PstI	9064 3564 1683	8741 3887 1683
Asp718I	10396 3915	12310 2001
HindII	4851 4137 2083 1022 923 727 566	4978 4851 1244 1022 923 727 566
BspEI	14311	

Figure 4.27. An RFLP study of four correctly orientated and one incorrectly orientated G394R pDB2536 transformant using five different restriction endonucleases. 1-2: Amp orientated, 3-5: LEU orientated and M: 1Kb Marker. Fragments highlighted in the blocked area correspond to the pUC9 region of pDB2536 following its disintegration from the vector.

4.2.4.2. Trehalose Stocks for Fermentation

After two days of growth, two colonies from each of the BMM-D pDB2536-hSTf plates and two from the BMM-D pSAC35 plate were used to inoculate 10mL of BMM-D broth. These cultures were incubated at 30°C for 2 days before being spun at 3000 rpm for 10 min. For each culture, 5mL of the supernatant was taken off and replaced with 5mL of 40% (v/v) D(+) trehalose (Sigma). Following resuspension of the pellet, 1mL aliquots were taken and frozen at -80°C in 1.8mL NUNC Cryotubes. One aliquot from each transformant was used to inoculate 100mL of LB broth to test for bacterial contamination. No contamination was observed.

4.2.4.3. Immunoprecipitation Agar Plate

To confirm that the cells were expressing hSTf, the same eight colonies picked to make trehalose stocks were also inoculated in small patches onto a BMM-D agar plate containing Anti-Human Transferrin serum (Calbiochem). After four days of incubation at 30°C, precipitation halos appeared around the transformants containing the hSTf contructs but not around the two containing just the pSAC35 (Figure 4.28).

4.2.5. Transferrin Expression

The conditions required for the production of cloned proteins by *S.cerevisiae* need to be very tightly controlled in order to optimise expression levels whilst maintaining cell growth.

To understand the reasoning behind the choice of conditions used, it is important to note that *S.cerevisiae* cells are known to exhibit a diauxic type of growth. Rather than the commonly seen growth pattern of lag phase, followed by exponential

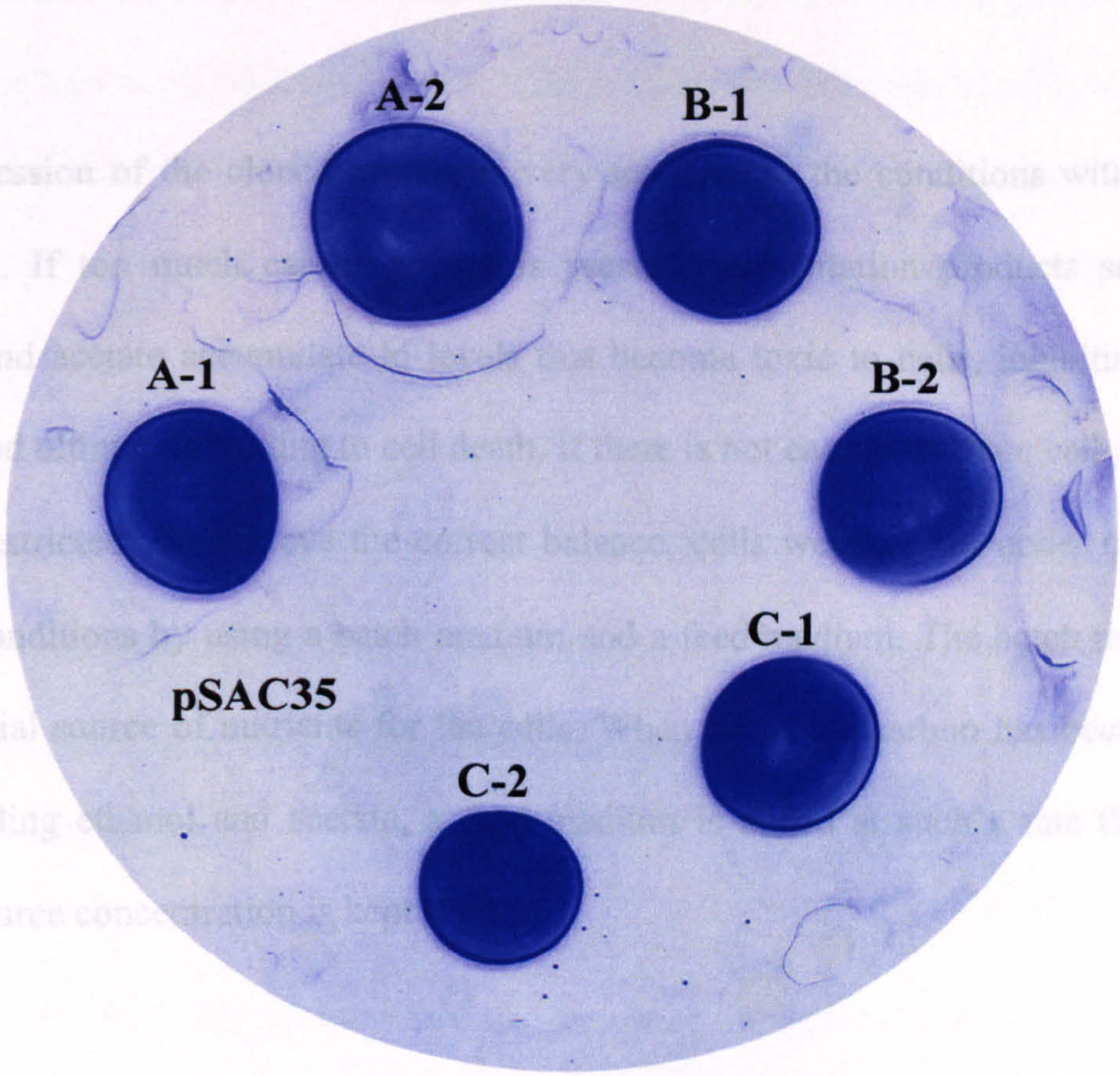


Figure 4.28. Immunodiffusion BMM-D agar plate. The plate was stained by firstly gently washing off the Yeast colonies off the agar with distilled water and then laying a GelBond® Film (BMA), hydrophilic-side down, onto the agar under a layer of filter paper and tissues. The next day, the film and agar were stained using the same protocol as for SDS-PAGE.

phase, stationary phase and then sporulation, or death, *S.cerevisiae* has been shown to exhibit an extra phase of growth separated from the first one by a short stationary phase. The reason for this is due to *S.cerevisiae's* ability to adapt to use the ethanol and acetate produced from the fermentation of sugars in the first exponential phase as a carbon source.

The expression of the cloned protein is very sensitive to the conditions within the fermenter. If too much carbon source is present, fermentation products such as ethanol and acetate accumulate to levels that become toxic to cells, inhibiting cell growth and ultimately leading to cell death. If there is not enough carbon, cell growth will be restricted. To achieve the correct balance, cells were grown under carbon-limited conditions by using a batch medium and a feed medium. The batch medium is the initial source of nutrients for the cells. When all of the carbon has been used up, including ethanol and acetate, a feed medium is added at such a rate that the carbon source concentration is kept limiting.

Yeast growth is also very sensitive to pH, with cells growing optimally at pH 5.5. The CO_2 produced following carbohydrate metabolism forms carbonic acid in solution and causes the pH to drop in the fermenter. Also, during acetate utilisation the pH increases. To ensure optimum growth, the fermenter has a feedback system to detect pH change, adding 17.5% ammonia to increase it, or 2M H_2SO_4 to decrease it, maintaining the pH at the optimum 5.5.

4.2.5.1. Fermentation

Before inoculation, the batch medium was prepared in a 10 L fermenter. 200mL of an autoclaved 50% (w/v) sucrose solution was added to 2560mL of distilled water and filtered, using a 0.22µm SuporCap™ sterile capsule, into the fermentor. To this, a mixture containing 50mL of a vitamin stock solution (6g/L calcium pantothenate, 10g/L nicotinic acid, 6g/L *m*-inositol, 1.5g/L thiamine-HCl and 30mg D-Biotin), 600mL of a salt mixture (28g/L KOH, 40g/L H_3PO_4, 7.17g/L $MgSO_4 \cdot 7H_20$ and 0.75g/L $CaCl_2 \cdot H_20$) and 40mL of distilled water was added following filtration through a 0.22µm Durapore filtration unit. A trace element solution (45mL containing 63.07g/L H_2SO_4, 3g/L $ZnSO_4 \cdot 7H_20$, 10g/L $FeSO_4.7H_2O$, 2.42g/L $MnSO_4 \cdot H_2O$, 79mg/L $CuSO_4 \cdot 5H_2O$, 0.5g/L $Na_2MoO_4 \cdot 2H_2O$ and 0.56g/L $CoCl_2 \cdot 6H_20$) was also added to the fermenter, again following filtration through a 0.22µm Durapore filtration unit. Both Durapore filtration steps included a flush-through with 125mL of distilled water.

Inoculation of the batch medium with *S. cerevisae* cells containing the respective construct was achieved by growing up the specific trehalose stock in 100mL of BMMS to approximately 1g of cells per litre and then adding approximately 60mg of these cells to the batch medium.

Following batch growth, the feed medium was added at a rate that maintained carbon limitation conditions. The feed medium comprised 5L of sterile 62.5% (w/v) sucrose containing a mixture of 120mL of vitamin stock, 250mL of salt solution and 30mL of distilled water, filtered using a 0.22µm Durapore filtration unit. The filter was flushed with 100mL of distilled water. The feed medium also contained a mixture of

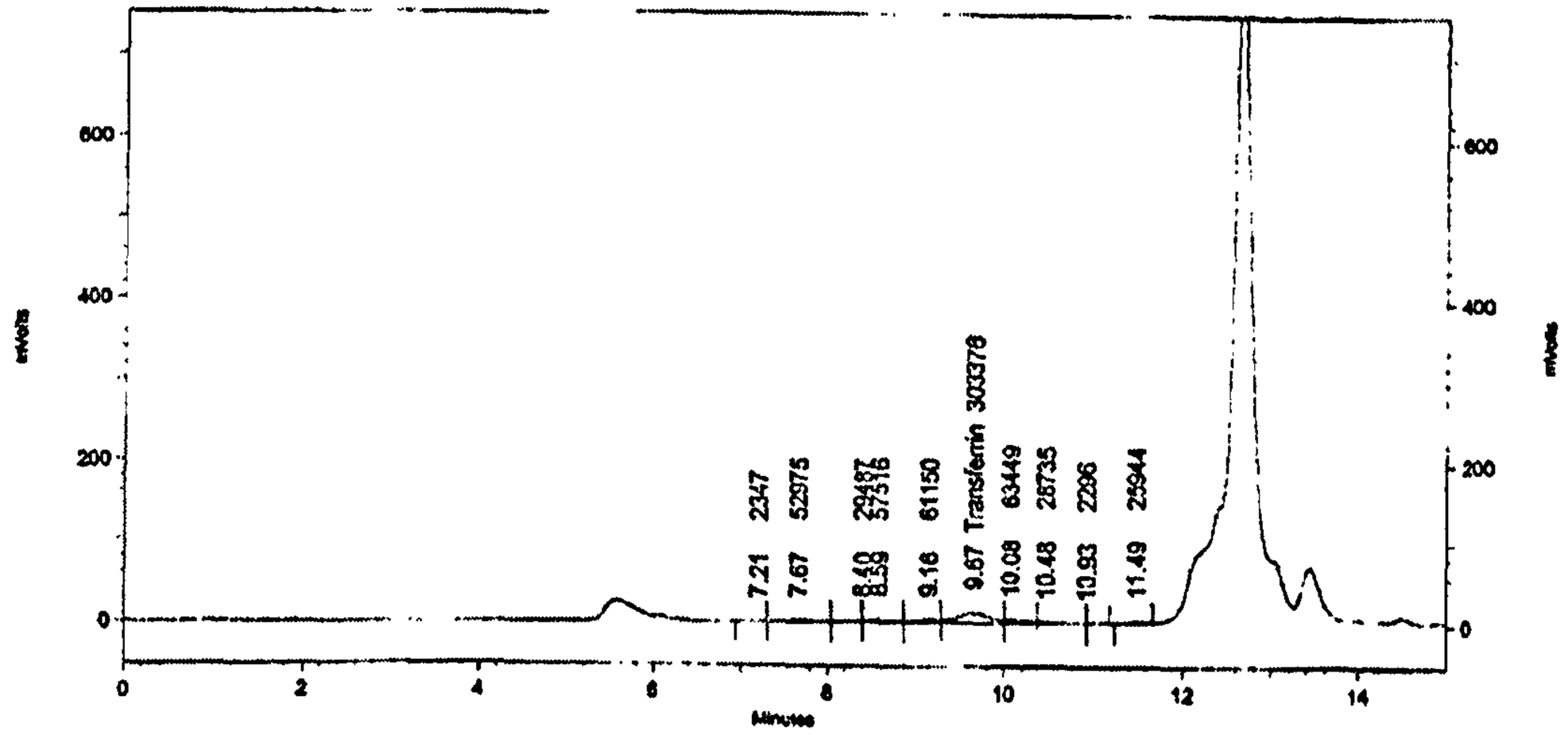
105mL of trace element solution and 295mL of distilled water, filtered and flushed in the same way as for the vitamin stock-salt solution. The fermentation was complete once the entire feed medium had been added.

In total, four fermentations were carried out. Two with *S.cerevisiae* cells containing the WT pDB2536-hSTf and two with *S.cerevisiae* cells containing the G394R mutated pDB2536-hSTf. Following fermentation, cultures were centrifuged at 5000rpm for 15 minutes and then the supernatant divided into 3 litre aliquots. These 3 litre aliquots were either processed straight away, or frozen at -20°C.

4.2.5.2. Protein Concentration Determination

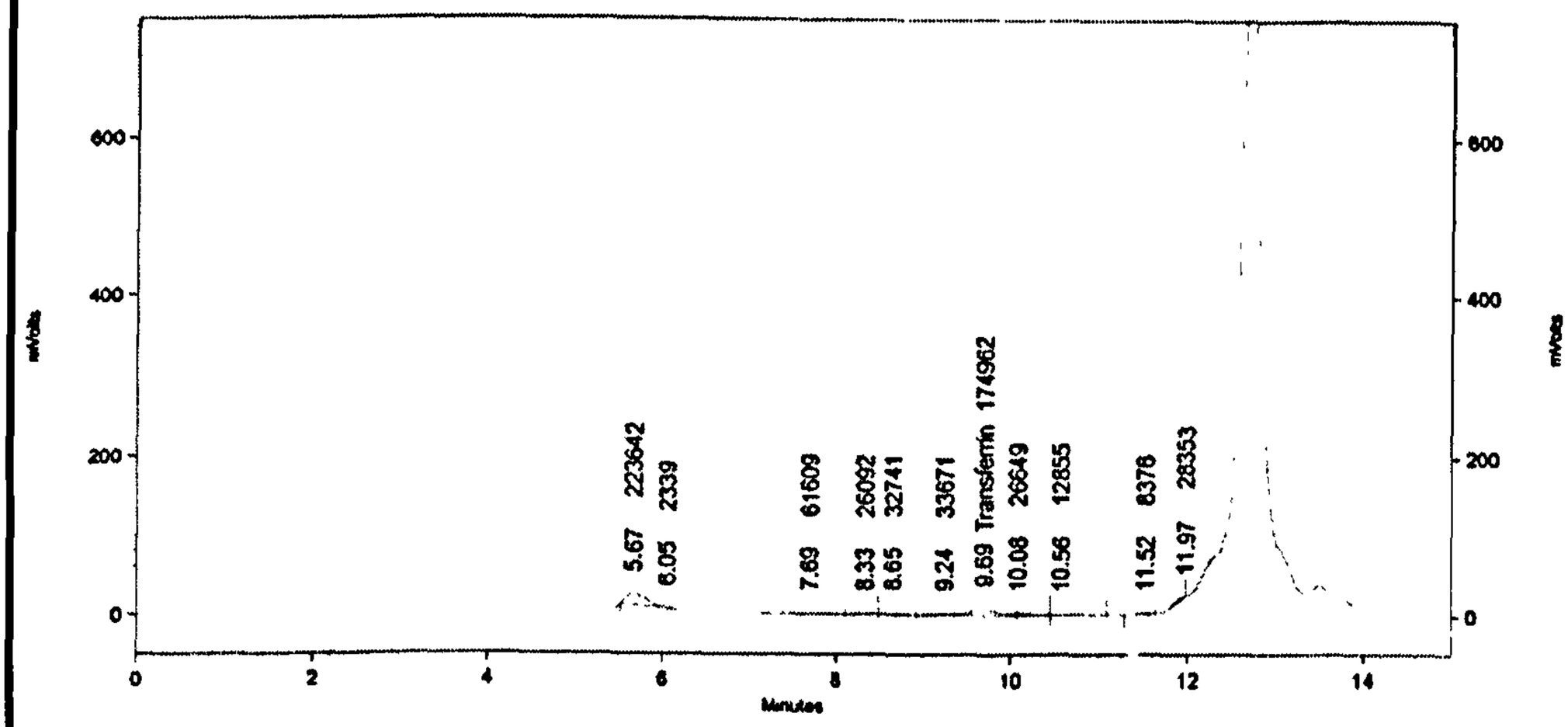
The protein concentration and relative purity of transferrin in the post-fermentation supernatant was determined by HPLC (section 2.4). Samples of supernatant (200µl) from both WT rhSTf and G394R rhSTf fermentations were loaded and run on the HPLC. The elution profile, shown in Figure 4.29, shows that the transferrin yield for WT rhSTf fermentation 1 was ~1.82mg/mL and for G394R rhSTf fermentation 1 was ~1.7mg/mL.

WT rhSTf Fermentation 1



Transferrin	9.67	303376	48.4	12026	1.82	mg/mL

G394R rhSTf Fermentation 1



Transferrin	9.69	174962	27.7	7409	1.7	mg/mL

Figure 4.29. The protein concentration and relative purity of transferrin in the post-fermentation cultures, determined by HPLC. The amount of rhSTf in mg/ml was adjusted to take into account dilution and sample volume.

5. Purification and Crystallisation

This chapter primarily focuses on the purification of the recombinant proteins (rhSTf) expressed in Chapter 4, along with the purification of native wild type (WT) hSTf, native G394R hSTf and native hLf. The production of highly pure preparations of these proteins is very important, since many of the structural and functional characterisation methods used later on in these studies were dependant on it.

As well as purification, this chapter also discuses the crystallisation of recombinant hSTf and native hLf-In^{3+} in the hope to solve their structures. The crystallisation of hSTf is of particular interest since only a high-resolution image of its N-lobe has so far been solved. As for hLf, even though this protein's structure has been solved at high-resolution, it has not been solved with indium bound to it.

5.1. Purification

5.1.1. Introduction

In order to characterise both the WT and G394R recombinant hSTf (rhSTf) and to compare them with both WT and G394R native hSTf, pure preparations of each of the four proteins is essential. Purification of both recombinant proteins was achieved using a number of chromatography steps. These steps are described in the following sections, along with the purification of both native proteins. Native G394R hSTf was purified from serum kindly donated to Dr R Evans by the individual who is heterozygous for this variant (section 1.3.9). Purification of this variant was a two-step process. The first step involved the isolation of hSTf from the serum and the

second step was the resolution of the variant from the WT. Native WT hSTf was bought from Scipac Ltd but further purified by immunoaffinity chromatography.

5.1.2. Recombinant hSTf Purification using Ion-Exchange Chromatography

The first step in the purification of WT rhSTf and G394R rhSTf involved the use of two separate ion-exchange chromatography columns. A SP-FF sepharose column was used first, followed by a DEAE-FF sepharose column. The difference between these two columns is the charge they carry at a specified pH. SP-FF sepharose is strongly cationic, whereas DEAE-FF sepharose is weakly anionic. The use of two differently charged matrices, one after the other, allowed rhSTf to be purified without the use of salt gradients.

5.1.2.1. SP-FF Purification

For each fermentation, only one 3 litre aliquot (section 4.2.5) was processed at a time. Prior to loading of this aliquot onto the SP-FF column, the supernatant was conditioned by firstly reducing its conductivity to ~3mS.cm^{-1} by the addition of MilliQ water and then lowering its pH to ~5 by the addition of 100% glacial acetic acid. Once conditioned, it was important to process the sample immediately so to reduce the chance of any protein degradation.

This conditioned supernatant was then purified using the protocol described in section 2.3.1, collecting the elution when the UV spectrophotometer recorded a peak at 280nm. 1mL samples were also taken from the load, the load flow-through, the wash flow-through, the elution and the 1M NaCl wash flow-through. The eluate was

collected, filtered through a Millipore-Durapore™ 0.22μm filter and stored at 4°C for a maximum of 2 days before being processed further.

The 1mL samples were run on a non-reducing SDS PAGE gel (section 2.4.12) to determine the efficiency of the column (Figure 5.1). This figure shows samples collected from four different SP-FF runs, two WT rhSTf fermentations and two G394R rhSTf fermentations. The four SP-FF elution samples were also analysed using HPLC (section 2.4.7) to determine relative protein purity and concentration (Figure 5.2 and 5.3).

5.1.2.2. DEAE-FF Purification

The SP-FF elute was further purified using a DEAE-FF column (section 2.3.1). Prior to loading the SP-FF eluate onto this column, it was first conditioned by bringing its conductivity down to ~3mS.cm^{-1} with MilliQ water and increasing its pH to ~9 with 10M NaOH. As soon as the DEAE-FF eluate had been collected it was filtered through a Millipore-Durapore™ 0.22μm filter and stored at 4°C for a maximum of 2 days before being processed further.

As with the SP-FF run, 1mL samples were taken from the load, the load flow-through, the wash flow-through, the eluate and the 1M NaCl wash flow-through from each DEAE-FF run. These samples were then visualised on a non-reducing SDS PAGE gel to determine the efficiency of the column (Figure 5.4). This figure shows samples collected from four different DEAE-FF runs, two different WT rhSTf fermentations and two different G394R rhSTf fermentations. As with the SP-FF elutes, these four DEAE-FF eluates were also analysed using HPLC

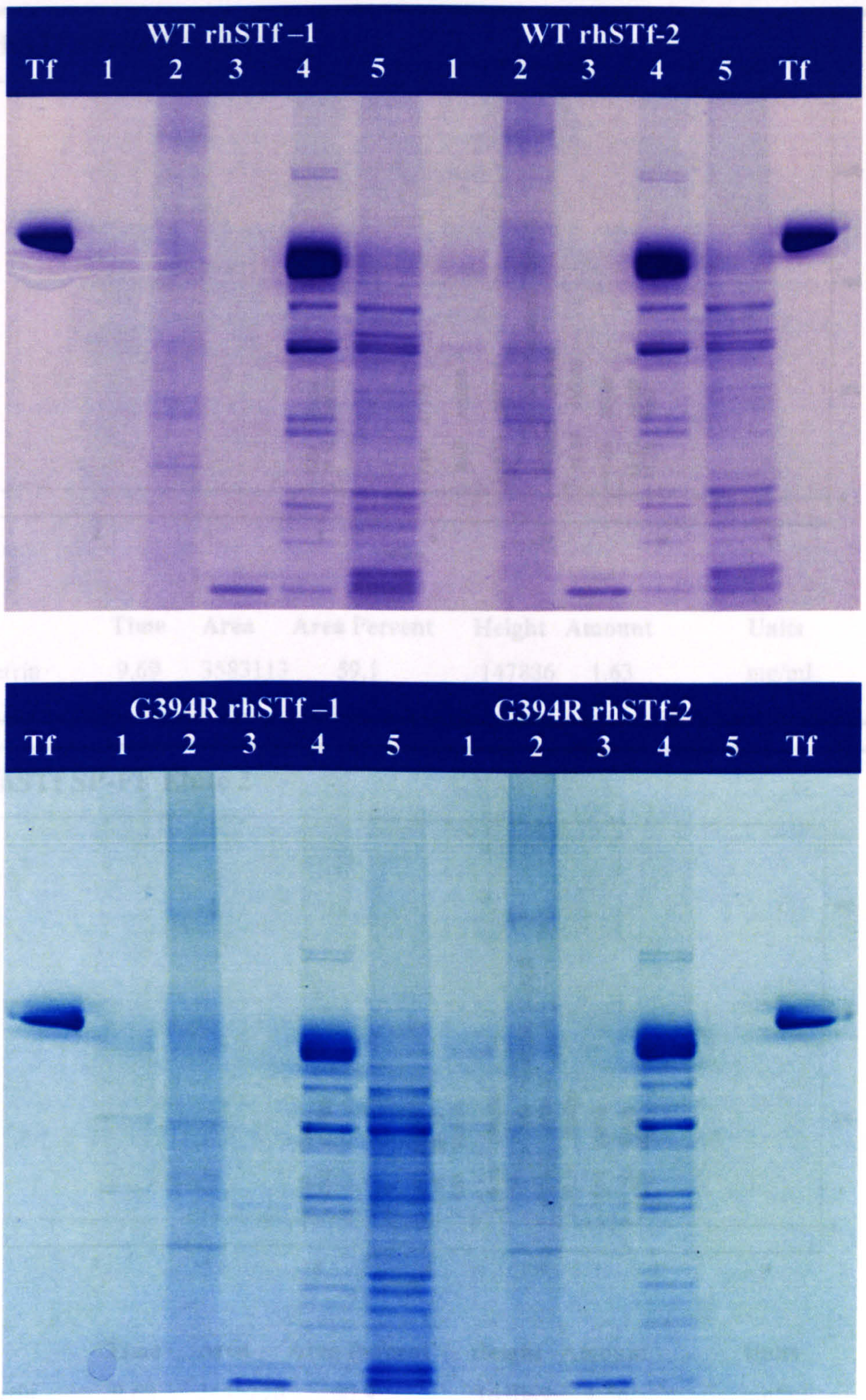
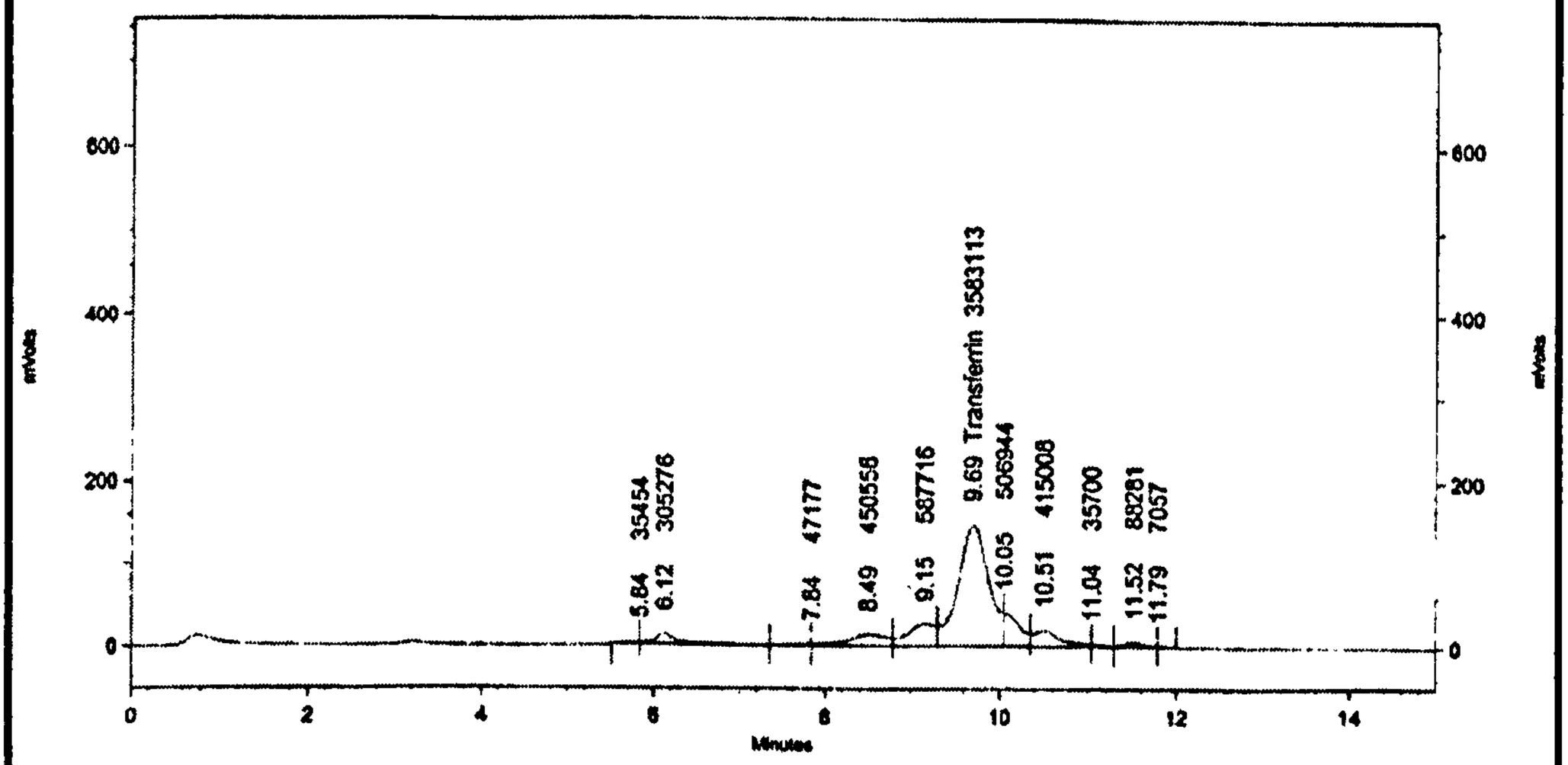


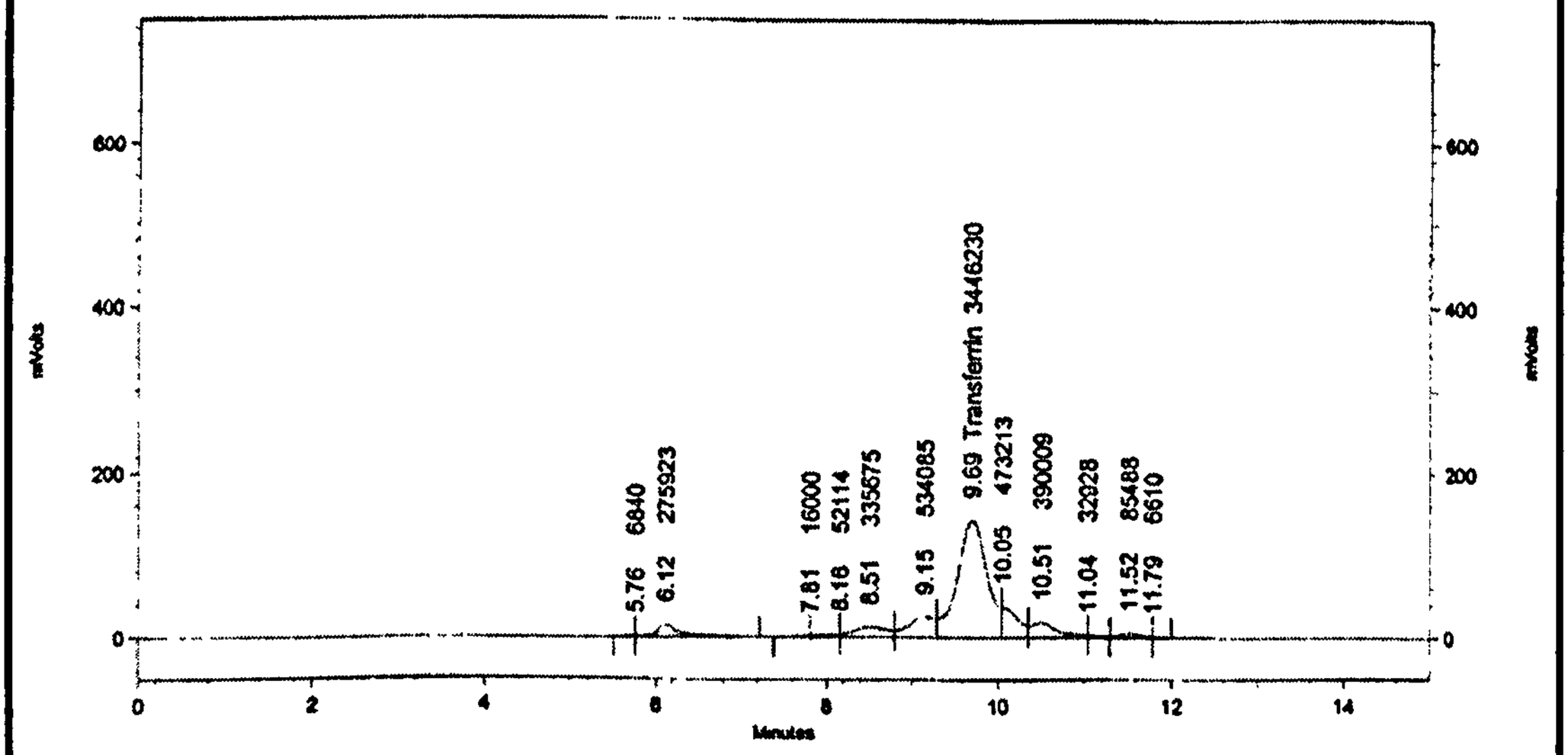
Figure 5.1. Non-reducing SDS PAGE of samples collected during SP-FF rhSTf purification. Tf: 1μg of native hSTf (Calbiochem), 1: load (diluted 1:10), 2: load flow-through, 3: wash flow through, 4: elution (50mM sodium dihydrogen phosphate di-hydrate, 100mM NaCl, pH 7) and 5: 1M NaCl wash flow-through

WT rhSTf SP-FF Elute 1



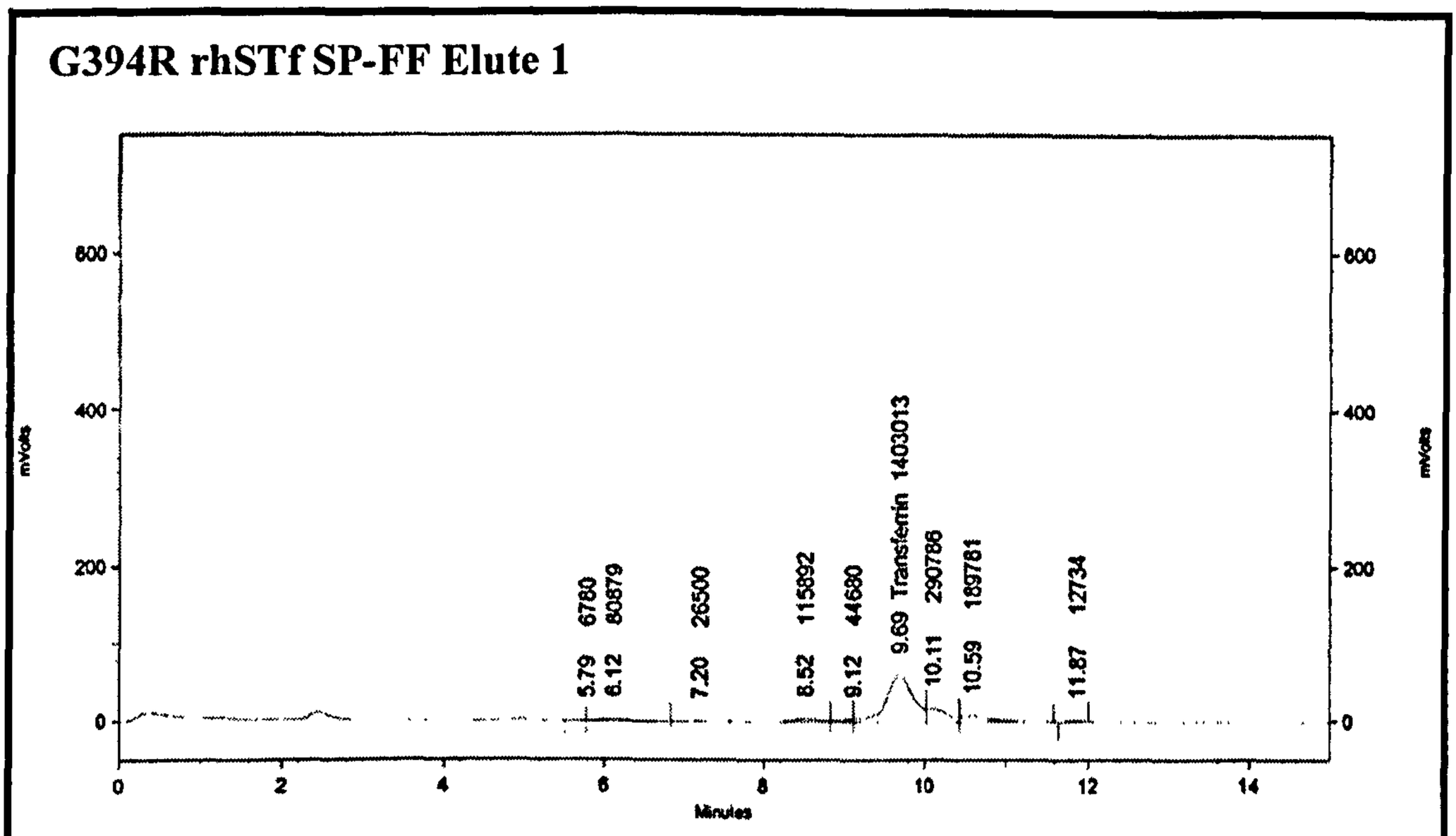
Name	Time	Area	Area Percent	Height	Amount	Units
Transferrin	9.69	3583113	59.1	147886	1.63	mg/mL

WT rhSTf SP-FF Elute 2

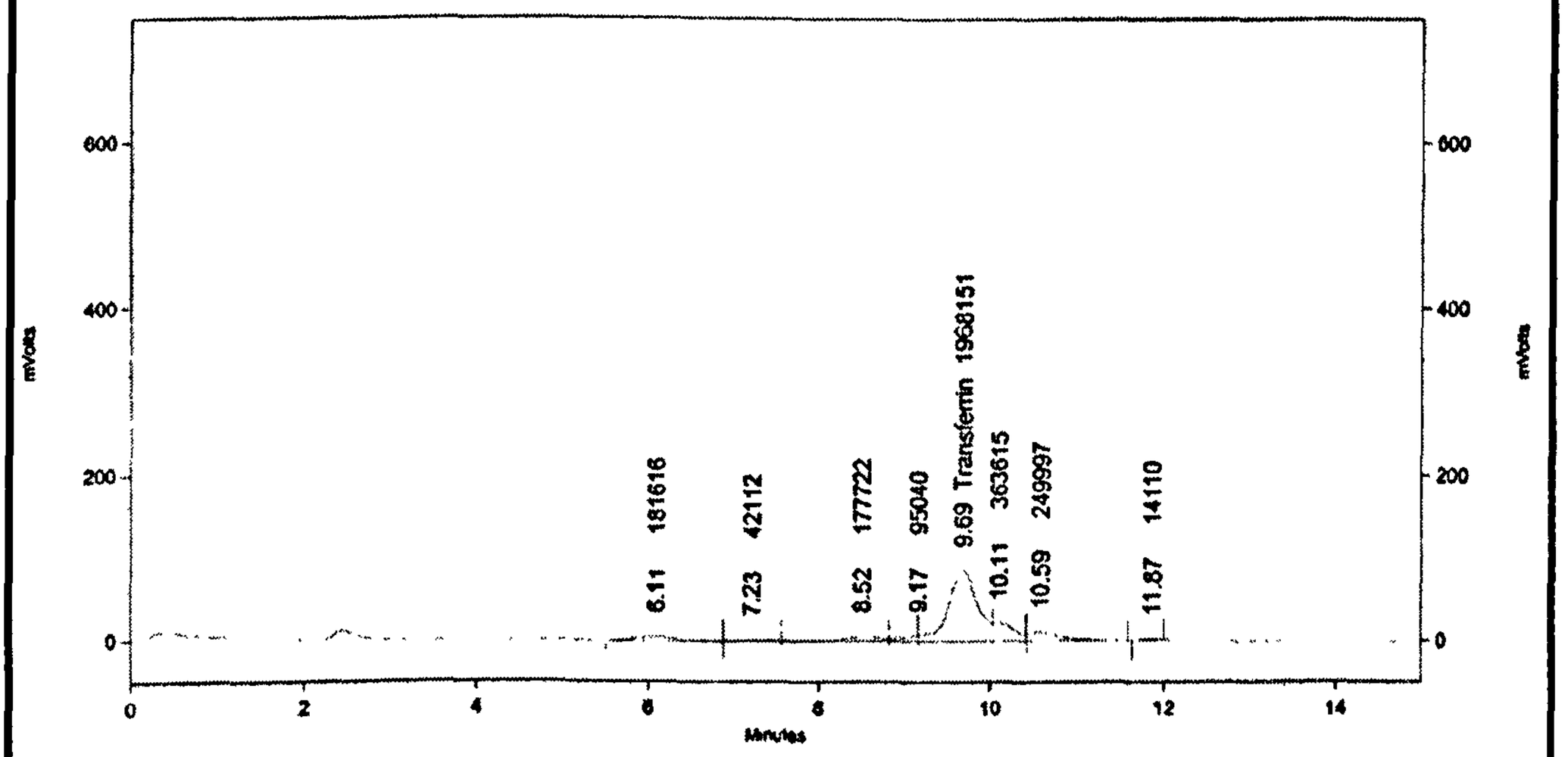


Name	Time	Area	Area Percent	Height	Amount	Units
Transferrin	9.69	3446230	60.9	143069	1.56	mg/mL

Figure 5.2. The protein concentration and relative purity of transferrin in the SP-FF eluate, determined by HPLC. The amount of rhSTf in mg/ml was adjusted to take into account dilution and sample volume.

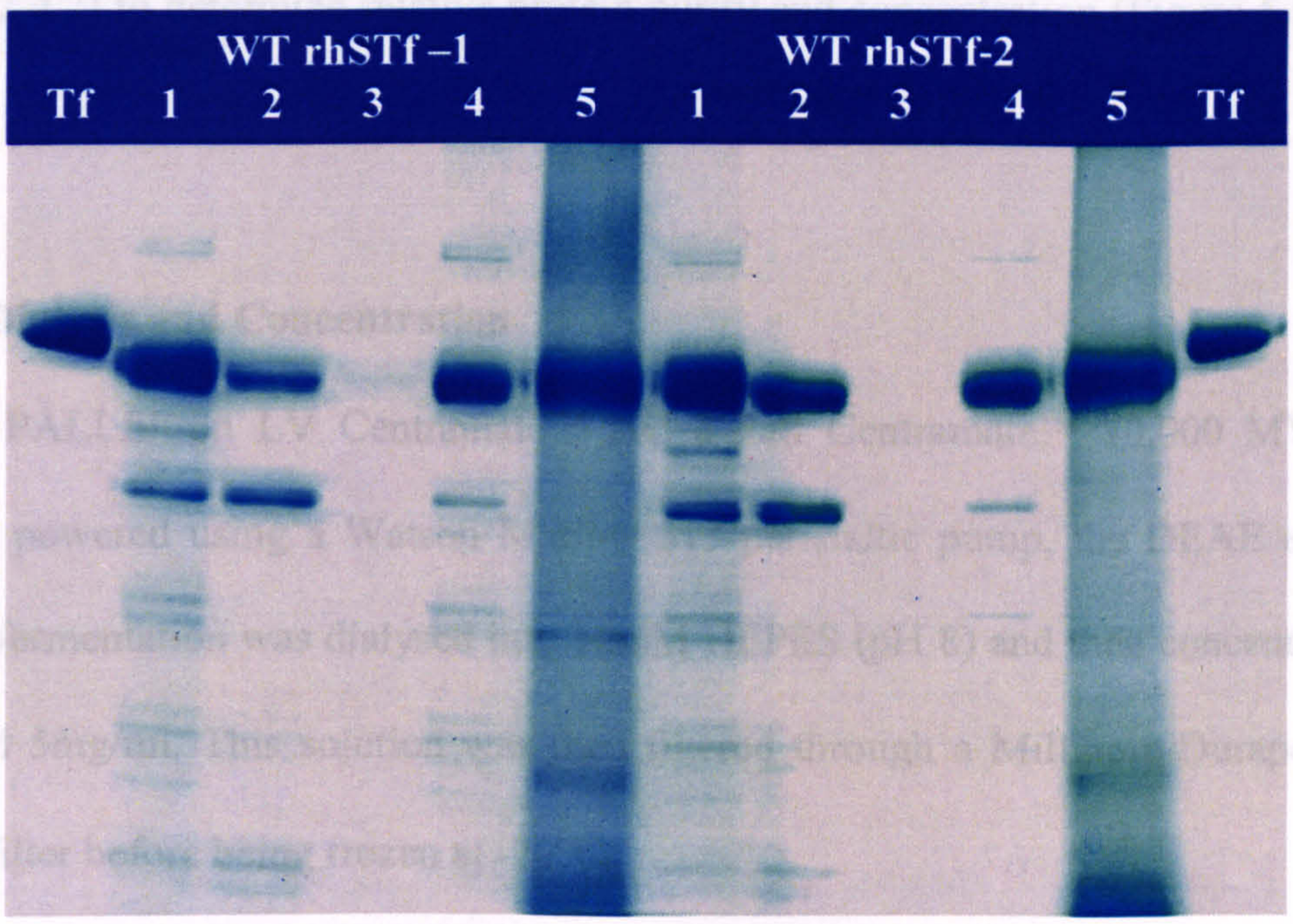


Name	Time	Area	Area Percent	Height	Amount	Units
Transferrin	9.69	1403013	64.6	61094	1.22	mg/mL



Name	Time	Area	Area Percent	Height	Amount	Units
Transferrin	9.69	1968151	63.6	85357	1.36	mg/mL

Figure 5.3. The protein concentration and relative purity of transferrin in the SP-FF eluate, determined by HPLC. The amount of rhSTf in mg/ml was adjusted to take into account dilution and sample volume.



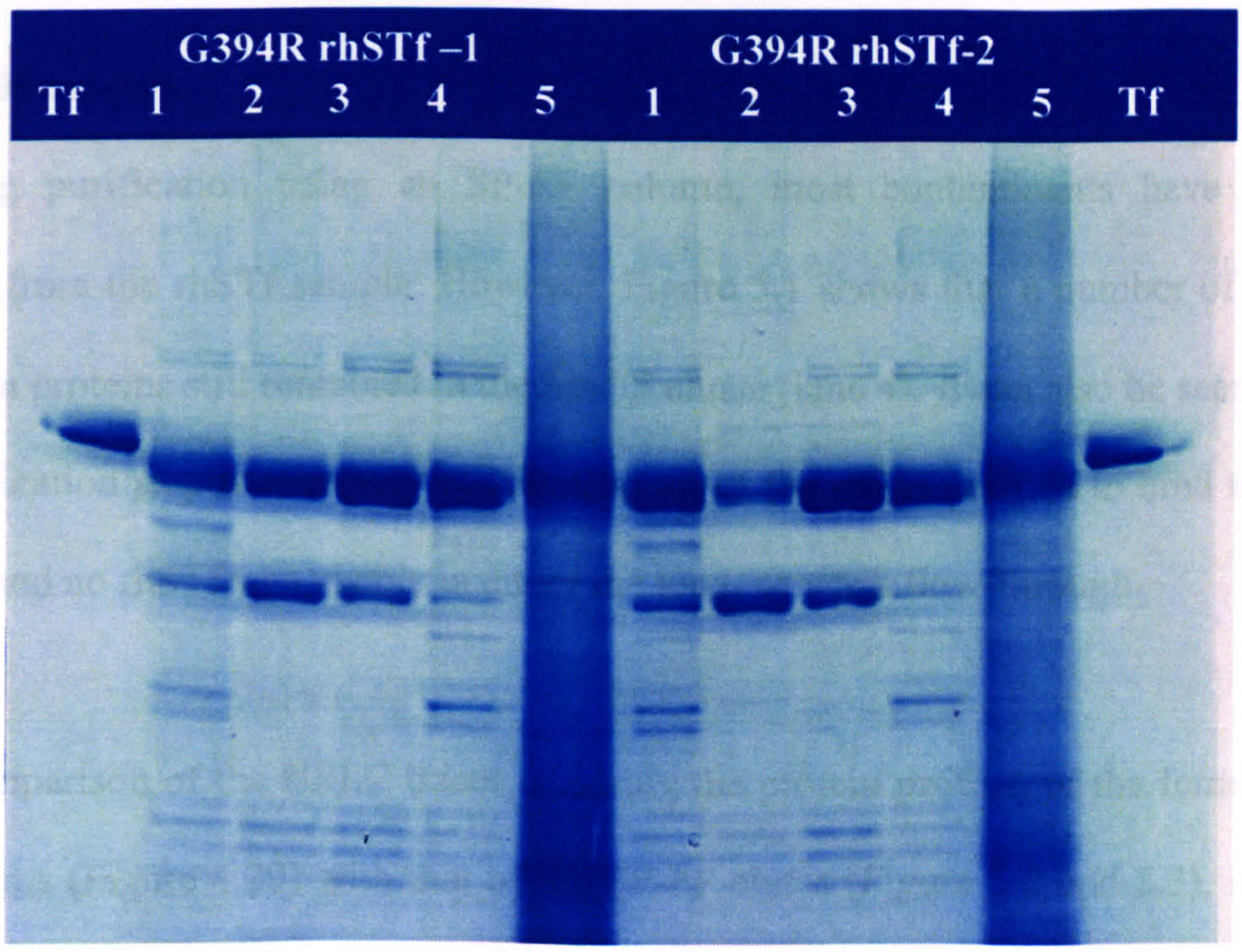


Figure 5.4. Non-reducing SDS PAGE of samples collected during DEAE-FF rhSTf purification. Tf: 1μg of native hSTf (Calbiochem), 1: load (diluted 1:10), 2: load flow-through, 3: wash flow through, 4: elution (82.6mM potassium tetraborate tetrahydrate, pH 9.4) (diluted 1:10) and 5: 1M NaCl wash flow-through

(section 2.4.7) to determine relative protein purity and concentration (Figure 5.5 and 5.6).

5.1.2.3. Dialysis and Concentration

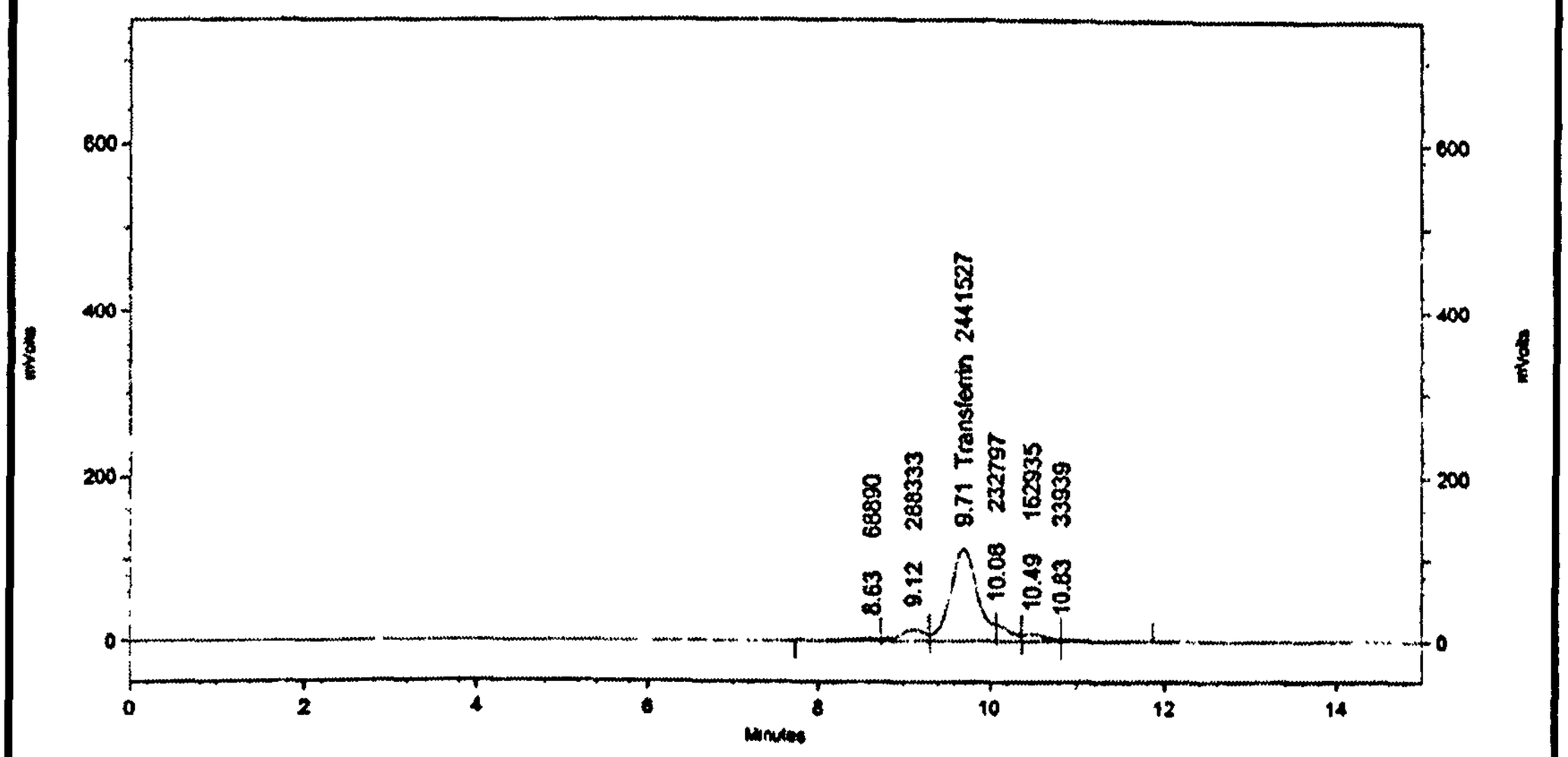
Using a PALLfiltron LV Centramate™ and a Pall Centramate™ 10,000 MWCO Cassette, powered using a Watson Marlow 313 peristaltic pump, the DEAE eluate for each fermentation was dialysed into 10mM HEPES (pH 8) and then concentrated to around 5mg/ml. This solution was then filtered through a Millipore-Durapore™ 0.22μm filter before being frozen at -20°C.

5.1.2.4. Discussion

Following purification using an SP-FF column, most contaminants have been removed from the rhSTf sample. However, Figure 5.1 shows that a number of non-transferrin proteins still remained in the SP-FF eluate (lane 4). It can also be seen that this purification step is relatively efficient, as all of the rhSTf seemed to bind to the column, and no rhSTf was visible in either the load- or wash flow-through.

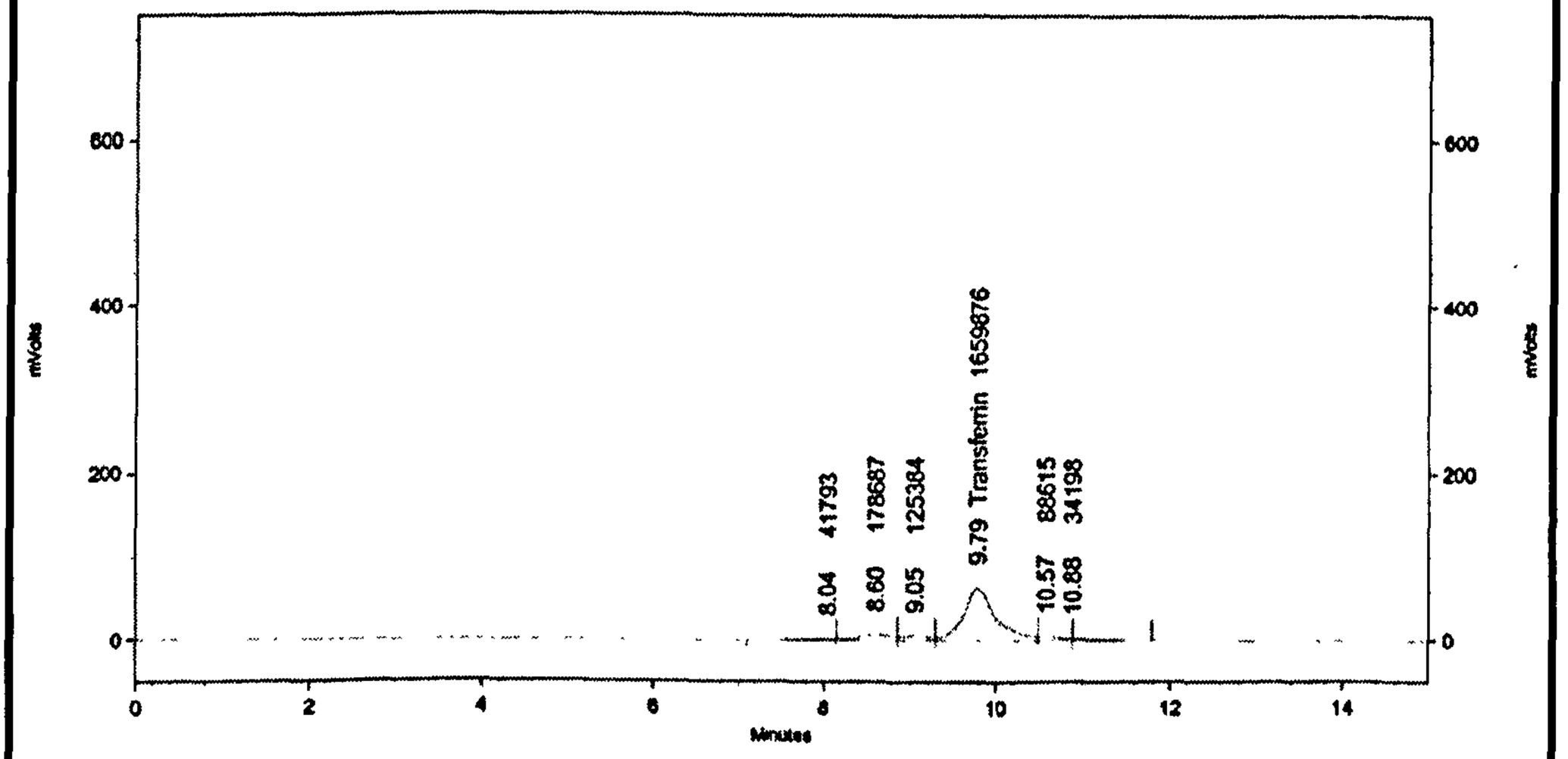
Upon comparison of the HPLC traces, showing the protein profiles of the fermenter supernatants (Figure 4.29) with that of the SP-FF elutes (Figure 5.2 and 5.3), it can be seen that a large proportion of the non-transferrin proteins have been removed. These traces also show roughly how much transferrin has been lost during the purification step by comparing the transferrin concentration (mg/mL) in the fermenter supernatant and the SP-FF elute, taking volume into account. Using these concentrations, the average percentage of protein recovered following SP-FF

WT rhSTf DEAE-FF Elute 1



Name	Time	Area	Area Percent	Height	Amount	Units
Transferrin	9.71	2441527	75.6	112497	1.21	mg/ml

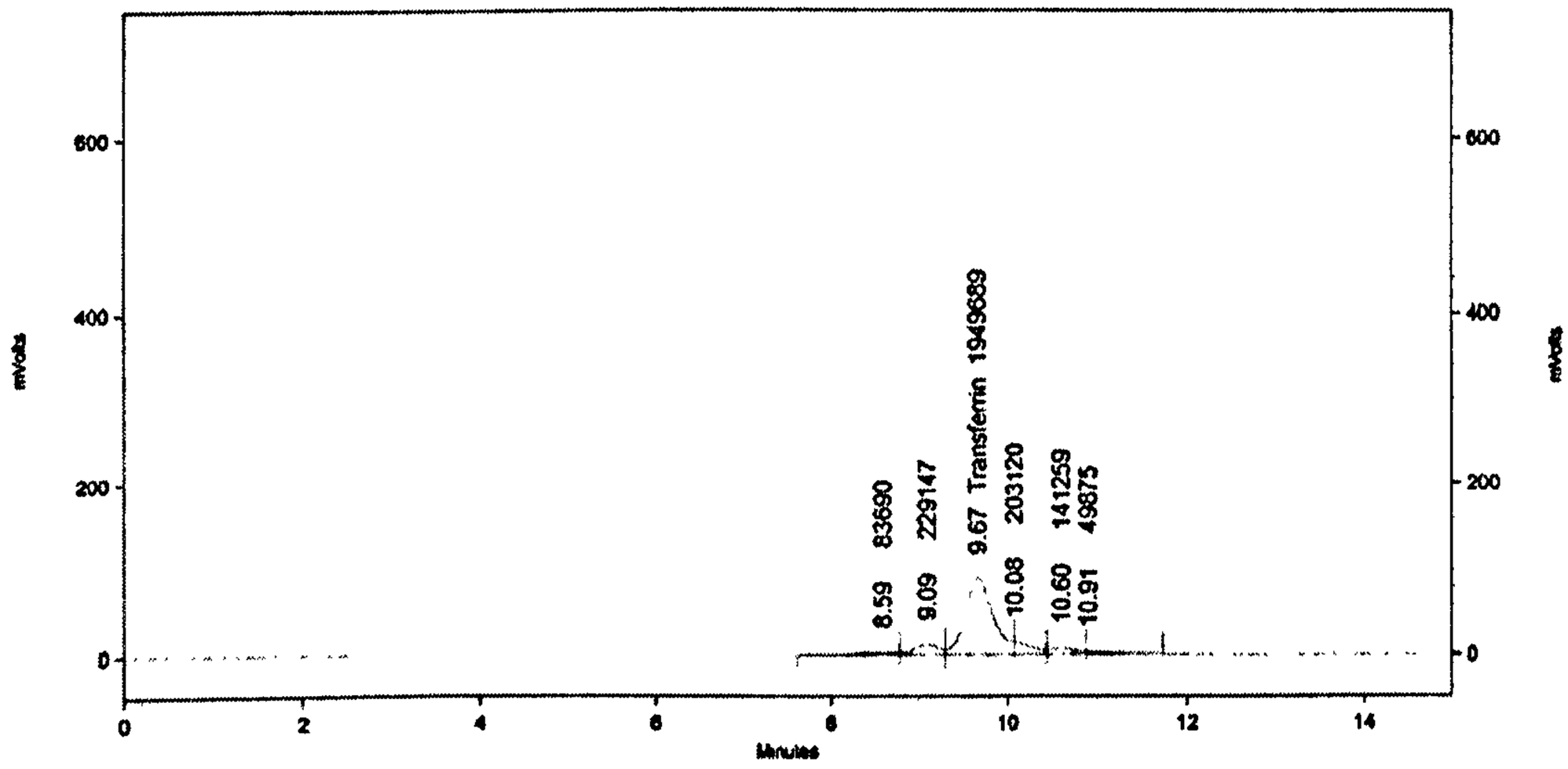
WT rhSTf DEAE-FF Elute 2



Name	Time	Area	Area Percent	Height	Amount	Units
Transferrin	9.79	1659876	78	63876	1.28	mg/ml

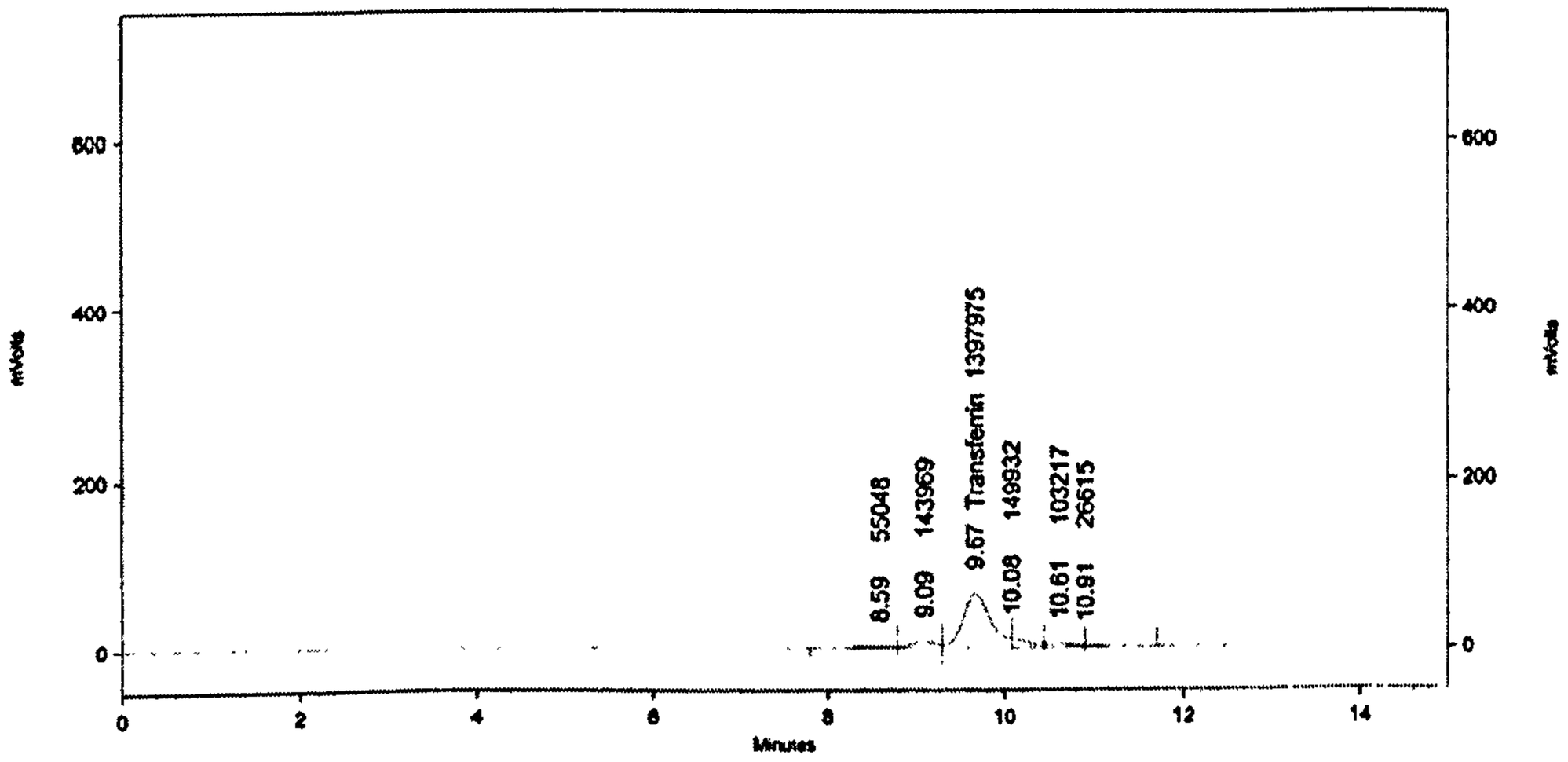
Figure 5.5. The protein concentration and relative purity of transferrin in the DEAE-FF elute, determined by HPLC. The amount of rhSTf in mg/ml was adjusted to take into account dilution and sample volume.

G394R rhSTf DEAE-FF Elute 1



Name	Time	Area	Area Percent	Height	Amount	Units
Transferrin	9.67	1949689	73.4	89257	1.02	mg/ml

G394R rhSTf DEAE-FF Elute 2



Name	Time	Area	Area Percent	Height	Amount	Units
Transferrin	9.67	1397975	74.5	64427	0.87	mg/ml

Figure 5.6. The protein concentration and relative purity of transferrin in the DEAE-FF elute, determined by HPLC. The amount of rhSTf in mg/ml was adjusted to take into account dilution and sample volume.

chromatography purification of the fermenter supernatant can be calculated. On average, 82% (88% for WT rhSTf and 76% for G394R rhSTf) of transferrin loaded onto the column was recovered in the eluate.

After the SP-FF eluate had been further purified on a DEAE-FF column, no real change in the purity of rhSTf was observed. Furthermore, a lot of rhSTf failed to bind to this column as large amounts were seen in the load flow-through (Figure 5.4). rhSTf was also seen in the wash flow-through of the G394R rhSTf DEAE-FF purification. The presence of rhSTf in both the load and wash flow-throughs indicates that in each case the DEAE-FF column was overloaded.

The protein profile seen in the HPLC charts (Figure 5.5 and 5.6) is also consistent with the SDS-PAGE results, in that a large portion of the rhSTf has been lost. This loss can be again calculated by comparing rhSTf concentration, taking into account volume, in the SP-FF eluate and the DEAE-FF eluate. For the two WT rhSTf purifications, the average protein recovery was approximately 78%. The recovery for the G394R rhSTf purification was even lower at 73%. This difference between the WT and G394R can also be seen on the SDS-PAGE gel in that substantially more protein seems to be present in the load and wash flow-through of the G394R purification than the WT purification. The average protein recovery using the DE-FF column can therefore be calculated to be 75.5%. This is 8% lower than that calculated for the SP-FF purification step.

5.1.3. Recombinant hSTf Purification using Immunoaffinity Chromatography

Following SP-FF and DEAE-FF chromatography, an immunoaffinity column was created to further purify rhSTf. The reason for creating such a column is so rhSTf can be purified efficiently, and without the need for gradients or complex buffers. This column was made using sheep anti-hSTf IgG coupled to cyanogen bromide activated Sepharose 4B (Amersham Biosciences). When the sample to be purified is run through this column, only rhSTf will bind to the IgG, leaving other proteins to be washed off. The rhSTf is then eluted off the column producing a very pure sample.

5.1.3.1. Creation of an Anti-hSTf Immunoaffinity Column

In order to make such a column, it was first necessary to purify sheep anti-hSTf IgG from the serum of sheep that have been immunised with hSTf. A very pure preparation of hSTf-specific IgG is important as if there were any other proteins present, they may interact with non-transferrin proteins and therefore reduce the efficiency of the column. To produce this pure preparation of IgG, a 50mL hSTf-coupled affinity column was first created using the protocol described in section 2.3.2 and 400mg of hSTf (Scipac). 40mL of sheep serum (Micropharm) was then loaded on to this column to purify IgG specific for hSTf. The purity of IgG in the eluate was checked using reducing and non-reducing SDS PAGE (section 2.4.12) (Figure 5.7). This process was repeated until approximately 400mg of IgG had been purified. The quantity of IgG was determined using UV spectroscopy and the same principals described in section 2.4.15, but using the extinction coefficient of 210,000$M^{-1}cm^{-1}$ and knowing the molecular weight of IgG is 150kDa. This 400mg of IgG was then used in the creation of a 50mL anti-hSTf immunoaffinity column using the same protocol for the creation of the hSTf-immunoaffinity column.

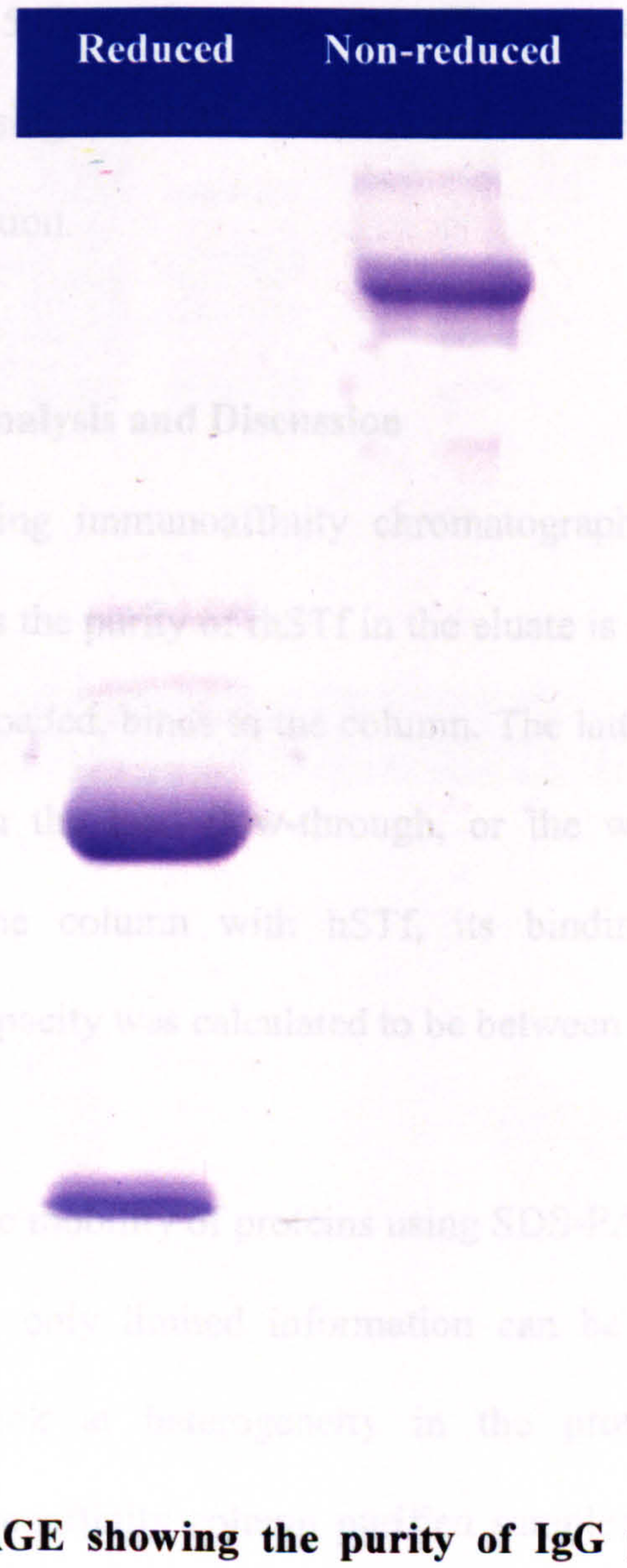


Figure 5.7. SDS PAGE showing the purity of IgG purified from sheep serum using an hSTf-immunoaffinity column. IgG eluted using 1M ammonia. By breaking the disulphide bridges using a reductant, the molecule's 2 light-chains (25KD) separate from the 2 heavy–chains (50KD). If IgG is not reduced, then the 4 chains remain bound to each other as a molecule with a molecular mass of 150kDa.

5.1.3.2. Purification of Recombinant hSTf

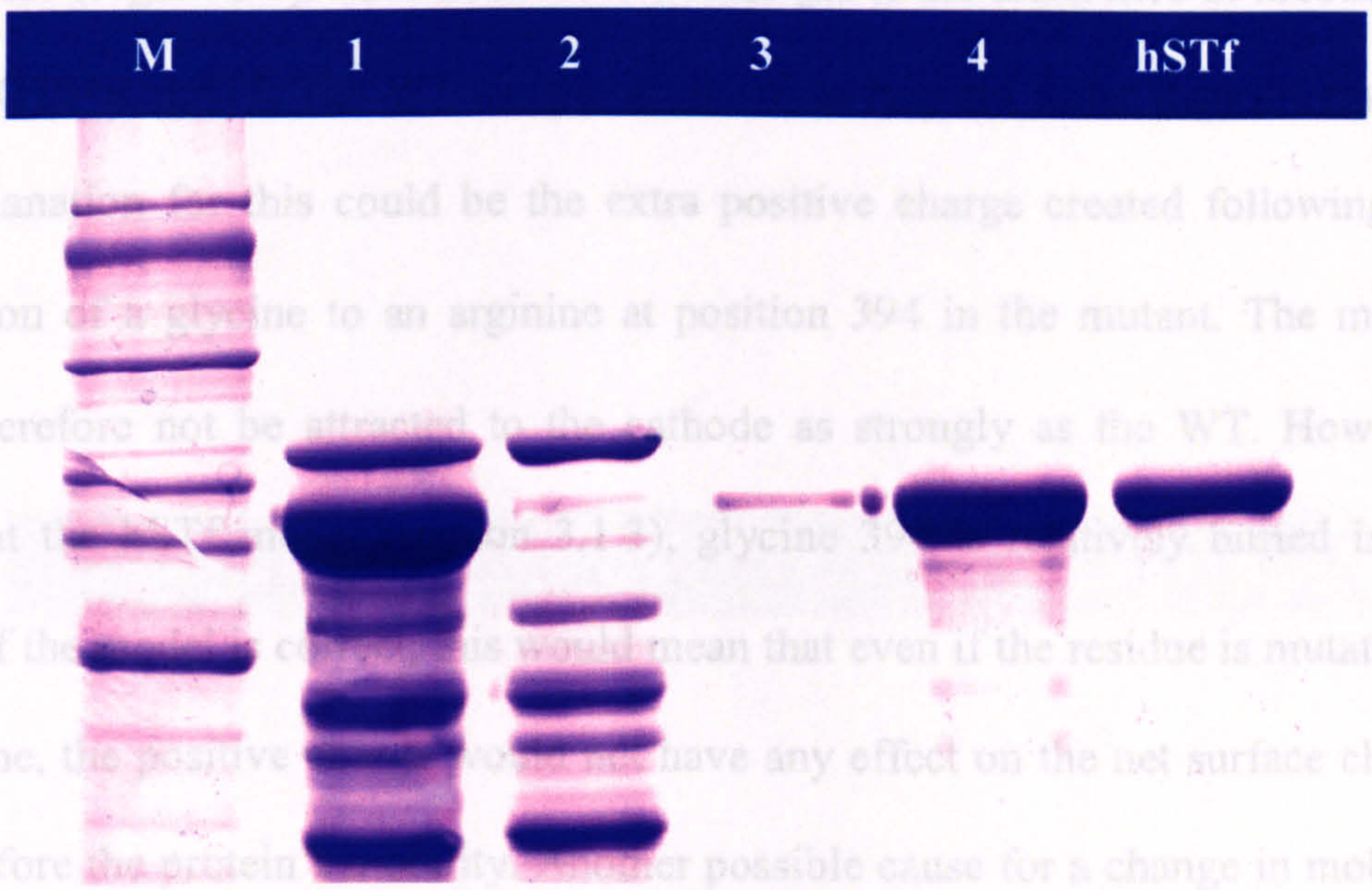
Using the protocol described in section 2.2.3, 40mL of each of the DEAE-FF eluates, in 10mM HEPES (pH 8 with 10M HCl), were run on the column. Samples were taken from the load, the load flow-through, the wash flow-through and the elution for each run. The protein profiles of each sample were visualised using SDS-PAGE (section 2.4.12) (Figure 5.8) to determine the efficiency of the column. Eluted protein was desalted using a G-25 column and freeze-dried (section 2.3.6) immediately after purification.

5.1.3.3. Native -PAGE Analysis and Discussion

Purification of rhSTf using immunoaffinity chromatography proved to be very efficient. Figure 5.8 shows the purity of rhSTf in the eluate is approaching 100% and that almost all the rhSTf loaded, binds to the column. The latter can be concluded as no rhSTf can be seen in the load flow-through, or the wash flow-through. By purposely overloading the column with hSTf, its binding capacity could be calculated. The binding capacity was calculated to be between 100 to 150mg.

Because the electrophoretic mobility of proteins using SDS-PAGE depends primarily on their molecular mass, only limited information can be obtained about these proteins. In order to look at heterogeneity in the protein's charge and its hydrodynamic size, immunoaffinity column purified samples of each protein were made diferric using the protocol described in section 2.3.5 and then run on native PAGE gels (section 2.4.10). This gives a rough indication of whether all of the proteins in the population have been correctly processed and whether there is any protein aggregation. SDS-PAGE shows the protein population of both rhSTf samples

A



B

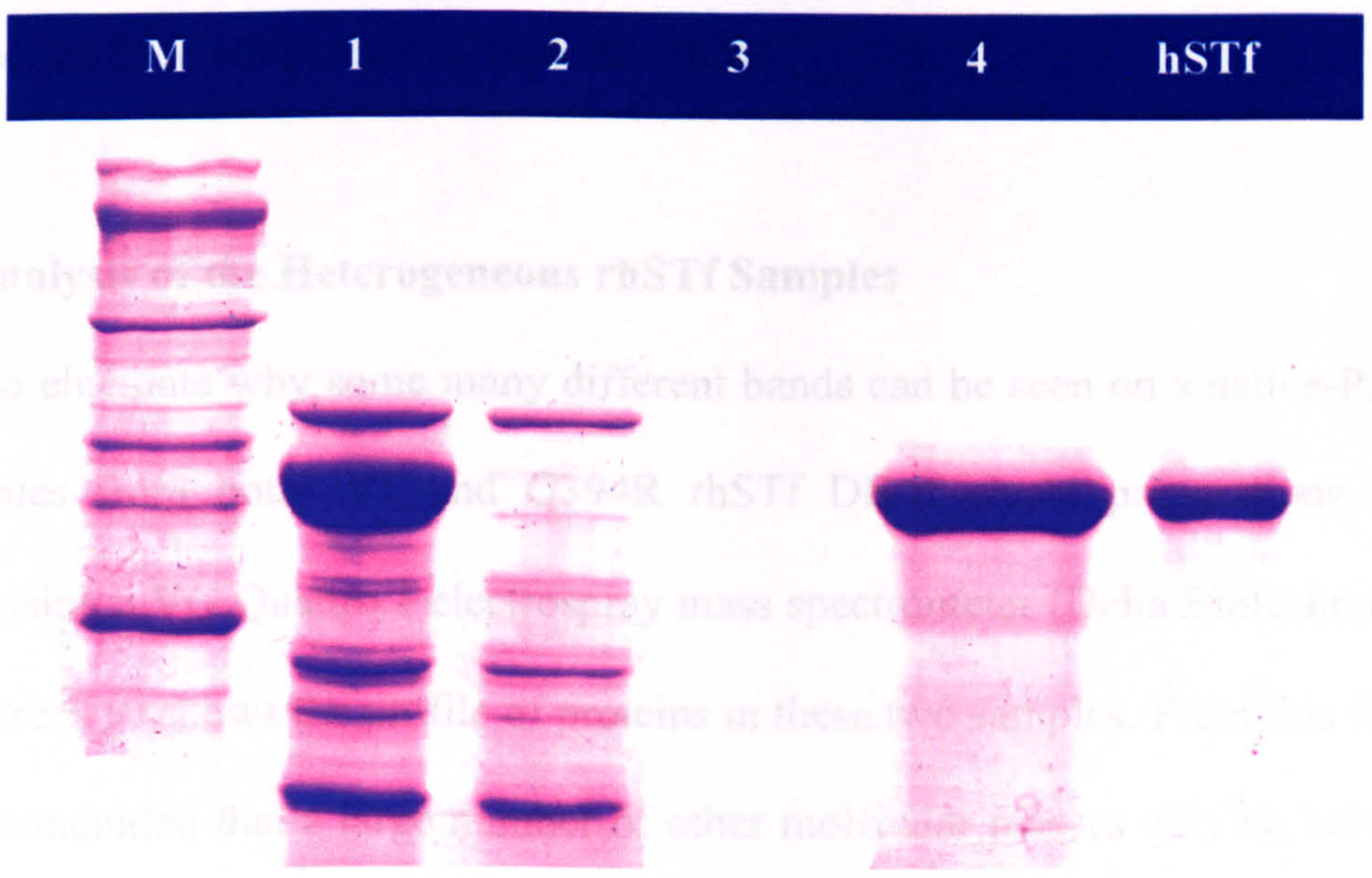


Figure 5.8. Reducing SDS-PAGE of samples collected during two rhSTf purifications using an anti-hSTf immunoaffinity column. Gel A: WT rhSTf and Gel B: G394R rhSTf. M: HMW-SDS marker, 1: Load, 2: Load flow-through, 3: Wash and 4: Eluate (eluted using 1M ammonia). hSTf: 1μg of native hSTf (Scipac Ltd)

to be homogeneous, whereas when using native-PAGE, a number of different forms were visible (Figure 5.9). Also noticeable on this gel is the difference in mobility of WT and G394R rhSTf. The WT sample is shown to migrate faster than the mutant. One explanation for this could be the extra positive charge created following the substitution of a glycine to an arginine at position 394 in the mutant. The mutant would therefore not be attracted to the cathode as strongly as the WT. However, looking at the hSTf model (section 3.1.3), glycine 394 is relatively buried in the protein. If the model is correct, this would mean that even if the residue is mutated to an arginine, the positive charge would not have any effect on the net surface charge and therefore the protein's mobility. Another possible cause for a change in mobility is a difference in the folding of the two proteins. As larger molecules migrate slower than smaller ones, it could be postulated that the mutation has possibly caused the C-lobe of the rhSTf to adopt a more open structure.

5.1.4. Analysis of the Heterogeneous rhSTf Samples

In order to elucidate why some many different bands can be seen on a native-PAGE gel, samples from both WT and G394R rhSTf DE-ff eluate preparations were analysed using a VG Quattro 1 electrospray mass spectrometer (Delta Biotechnology Ltd). Figure 5.10 shows the profile of proteins in these two samples. From this figure it can be concluded that a large number of other molecular masses varying between 75kDa and 77kDa are present in both samples. This variety is most likely to be due to different forms of rhSTf, rather than many different non-transferrin proteins. The reason for this is that the chances of so many proteins with a similar molecular mass still being present in the sample following SP-FF and DEAE-FF chromatography is

Figure 5.9. 8% Native PAGE of WT and G394R rhSTf following affinity chromatography. 1: WT rhSTf (4μg), 2: G394R rhSTf (4μg), 3: WT rhSTf (2μg) and 4: G394R rhSTf (2μg).

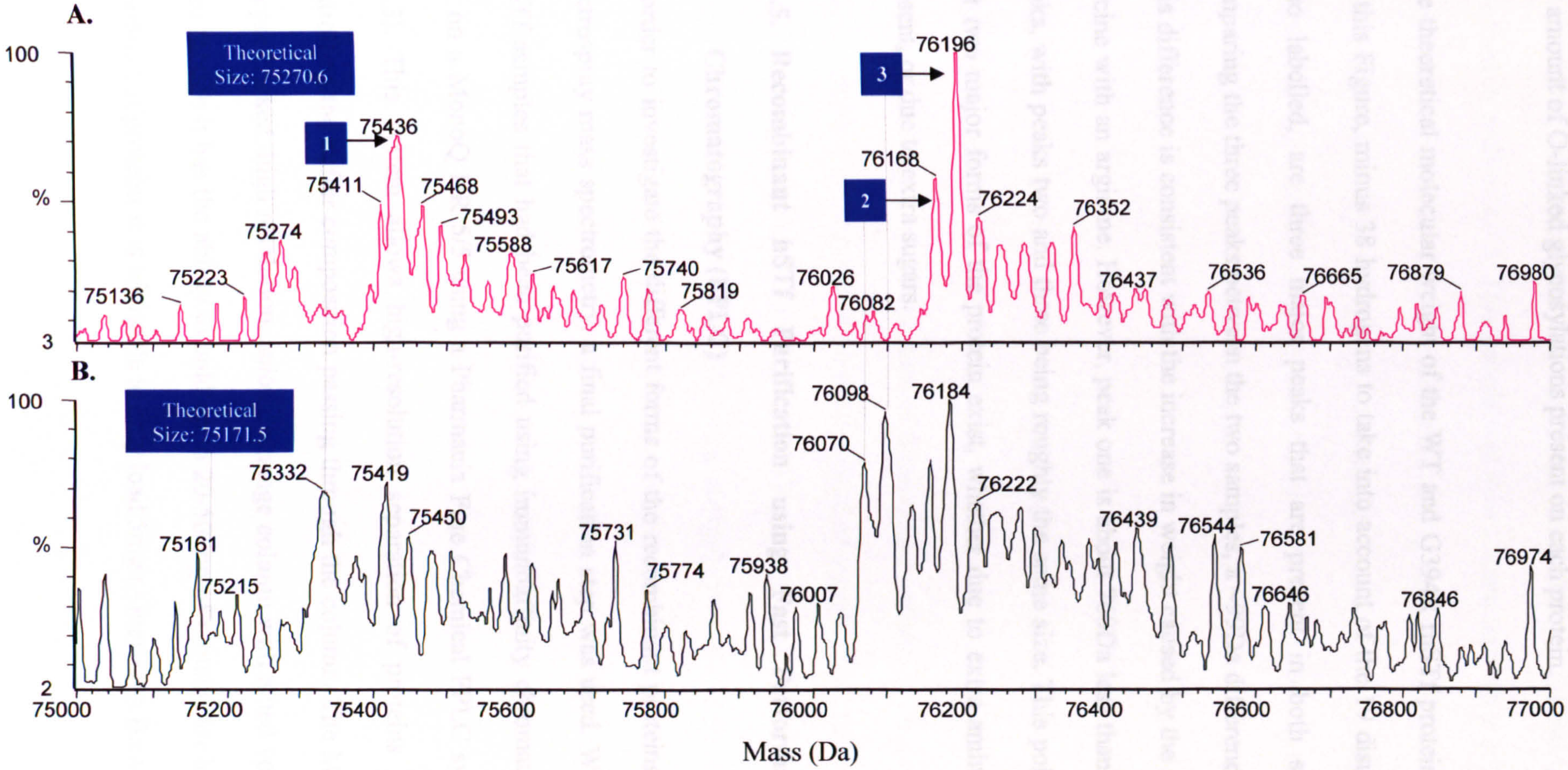


Figure 5.10. Analysis of WT and G394R rhSTf DE-FF eluate using a VG Quattro 1 electrospray mass spectrometer.

A: G394R rhSTf, B: WT rhSTf and 1-3: peaks that differ from each other by 99Da (±5Da) that may be due to the substitution of glycine (75.07Da) with arginine (174.2Da) at position 394.

very low. A possible explanation for this heterogeneity could be due to variations in the amount of O-linked glycosylations present on each protein.

The theoretical molecular weight of the WT and G394R rhSTf proteins are labelled on this Figure, minus 38 hydrogens to take into account of the 19 disulphide bonds. Also labelled, are three major peaks that are present in both spectra. When comparing the three peaks between the two samples, a ~99Da difference can be seen. This difference is consistent with the increase in weight caused by the substitution of glycine with an arginine. However, peak one is about 760Da less than the other two peaks, with peaks two and three being roughly the same size. This points to the fact that two major forms of the protein exist, whether due to extra amino acids being present, or due to extra sugars.

5.1.5. Recombinant hSTf Purification using Fast Performance Liquid Chromatography (FPLC)

In order to investigate the different forms of the recombinant proteins identified by electrospray mass spectrometry, a final purification step was used. WT and G394R rhSTf samples that had been purified using immunoaffinity chromatography were run on a MonoQ HR 5/5 using a Pharmacia Fine Chemical FPLC system (section 2.3.3). This system allows high-resolution separation of proteins by accurately controlling the buffer composition passing through the column. The MonoQ HR 5/5 is a pre-packed, high resolution, anion exchange column with a bed volume of 1 ml. Even though it has the ability to hold up to 20-50 mg of protein, for high-resolution separation of proteins it is advisable to only load 5mg (Amersham Biosciences).

5.1.5.1. Purification 1

To start, 5mg of freeze-dried WT rhSTf was made diferric following the protocol described in section 2.3.5. This sample was then filtered through a 0.2μ filter before being injected onto the 2mL loading-loop. Using program 1 (Table 5.1), a 2mL sample was loaded onto the column, washed and then eluted using a gradient of 0 - 250mM NaCl (Figure 5.11a). Samples from nine fractions spanning over the elution peak were collected and analysed using native PAGE (Figure 5.11b) as described in section 5.1.3.3.

From these two figures, it can be concluded that only fraction 1 contained a homogeneous sample of rhSTf. All other fractions still contained heterogeneous samples. Following closer inspection of these nine fractions, it can be seen that the band in sample 1 corresponds with the top band of sample 2 and the bottom band in sample 2 corresponds to the bottom band in sample 3. Also present in sample 3, is another band that migrates halfway between the other two. These three bands appear to correspond to the first three peaks on the elution trace. In the remaining six fractions, a fourth band can be visualised lower down than any of the other three. Because no other bands are clearly visible, it can be concluded that approximately four main forms of the protein exist in the rhSTf preparation.

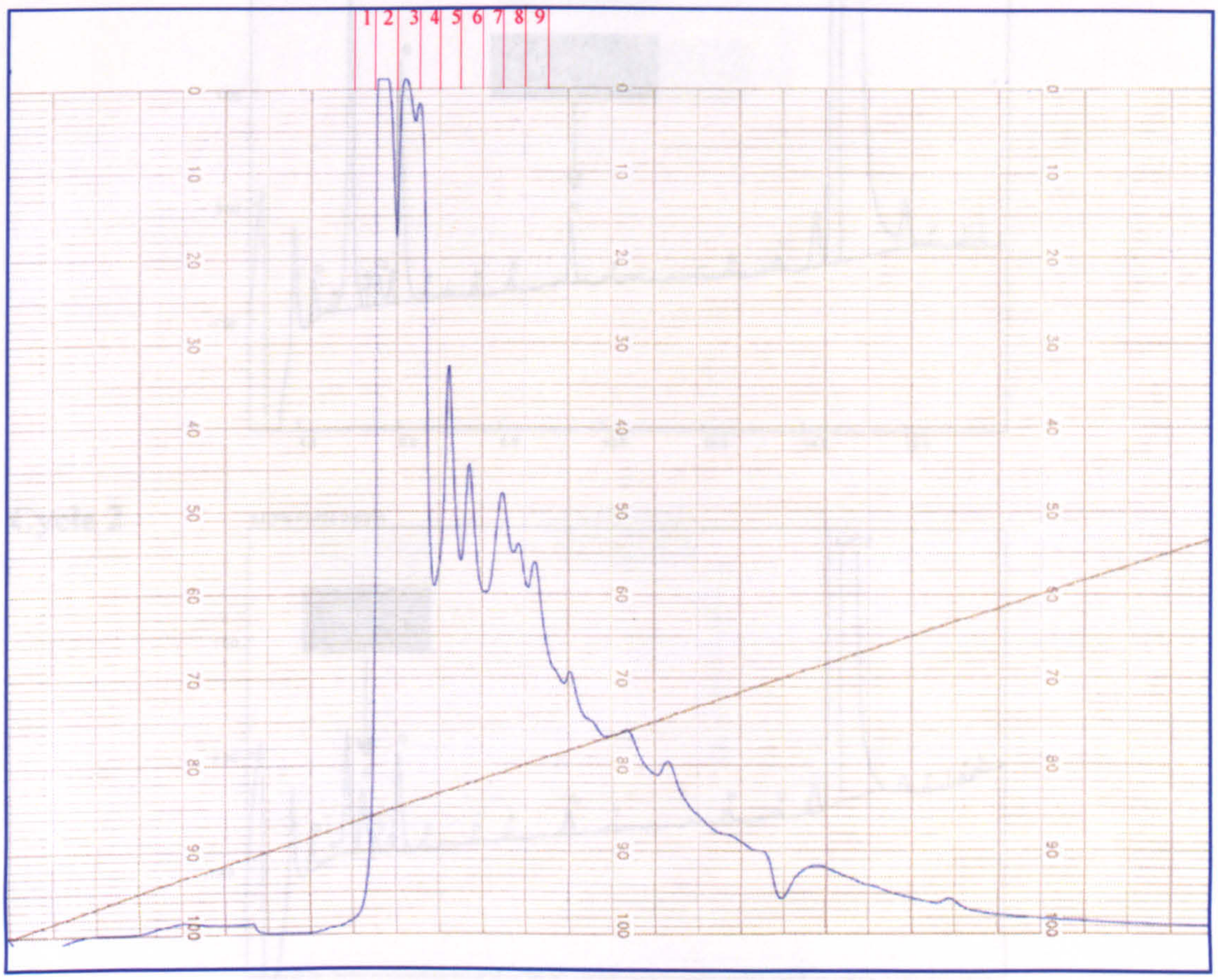
5.1.5.2. N-Terminal Sequencing

In order to elucidate the first two peaks, the protein in fraction 1 and 2 were N-terminal sequenced (Delta Biotechnology Ltd). N-terminal sequence data for both of these fractions are shown in Figure 5.12 and 5.13. This data reveals that in fraction one, only one protein exists that has an N-terminus that starts with arginine, followed

Program	Time (minutes)	Gradient Programmer GP-250 Command
1	0	Set port 3.0 and 6.0
		Buffer B = 0%
	5	Buffer B = 0%
		Set port 3.1 and 6.1
	125	Buffer B = **50%**
		Set port 3.0 and 6.0
		Buffer B = 100%
	130	Buffer B = 100%
		Buffer B = 0%
	135	Buffer B = 0%
2	0	Set port 3.0 and 6.0
		Buffer B = 0%
	5	Buffer B = 0%
		Set port 3.1 and 6.1
	125	Buffer B = **25%**
		Set port 3.0 and 6.0
		Buffer B = 100%
	130	Buffer B = 100%
		Buffer B = 0%
	135	Buffer B = 0%
3	0	Set port 3.0 and 6.0
		Buffer B = 0%
	5	Buffer B = 0%
		Set port 3.1 and 6.1
	125	Buffer B = **25% (0.4M NaCl)**
		Set port 3.0 and 6.0
		Buffer B = 100%
	130	Buffer B = 100%
		Buffer B = 0%
	135	Buffer B = 0%
4	0	Set port 3.0 and 6.0
		Buffer B = 0%
	5	Buffer B = 0%
		Set port 3.1 and 6.1
	145	Buffer B = **40%**
		Set port 3.0 and 6.0
		Buffer B = 100%
	150	Buffer B = 100%
		Buffer B = 0%
	155	Buffer B = 0%

Table 5.1. The programs executed for the purification of rhSTf using a GP-250 controlled FPLC system and a 1ml MonoQ HR 5/5 column. At the start of each program the flow rate is set to 0.5ml/min and the chart speed to 0.25cm/min. Fractions were collected every 2 minutes. Buffer A: 20mM Tris (pH8), buffer B: 20mM Tris, 0.5M NaCl (pH8) (except for program 3 where NaCl is at 0.4M), port 3: fraction collector and port 6: Valve V-7.

a)



b)

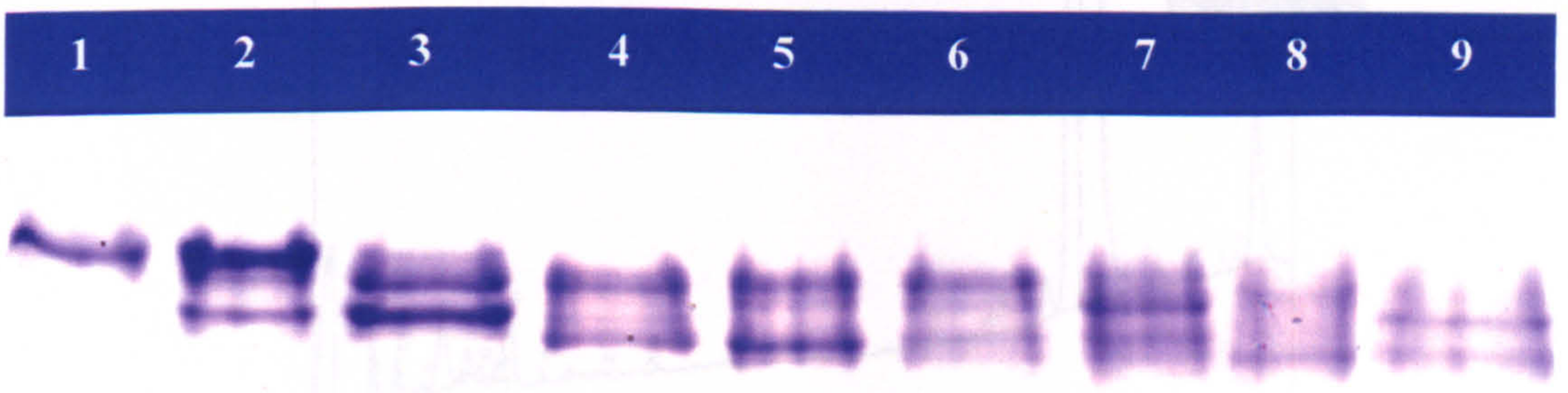
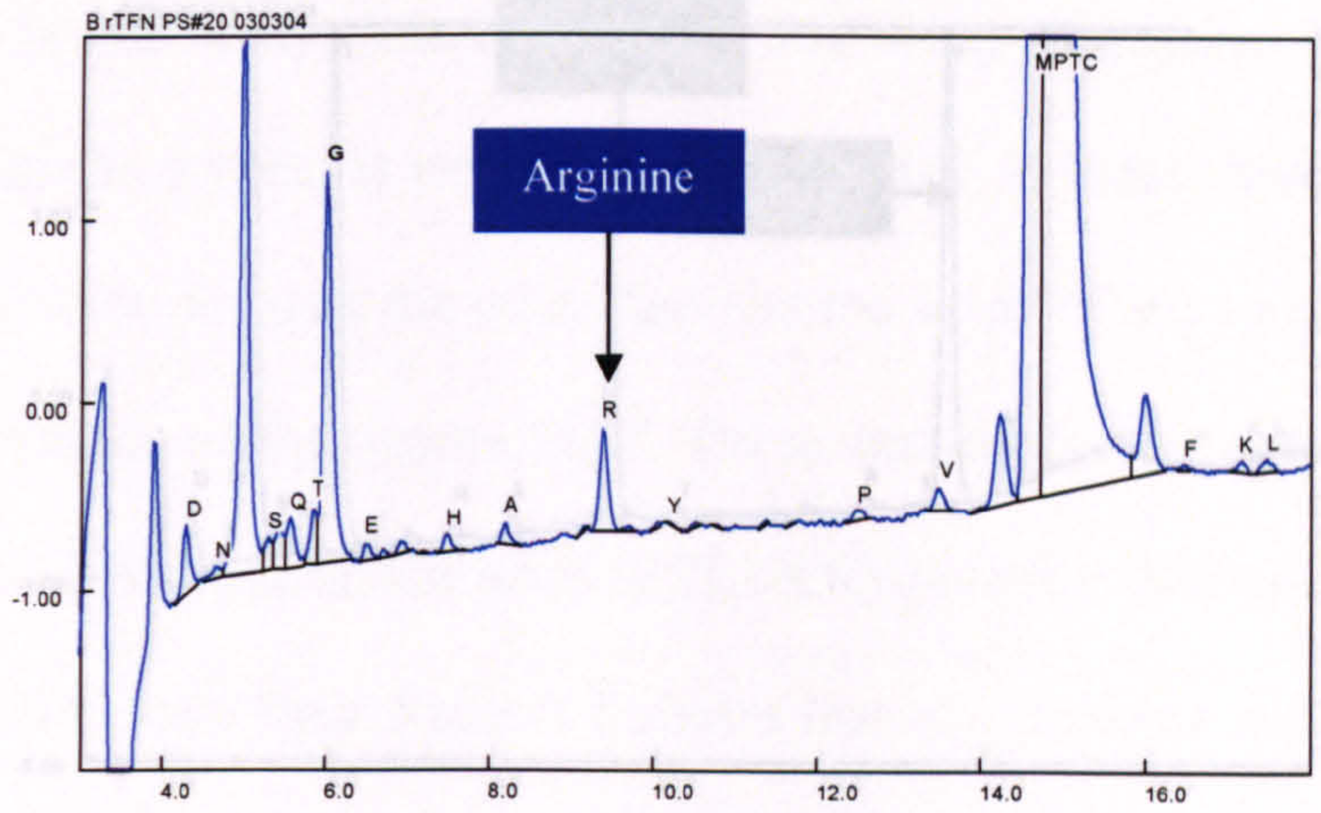
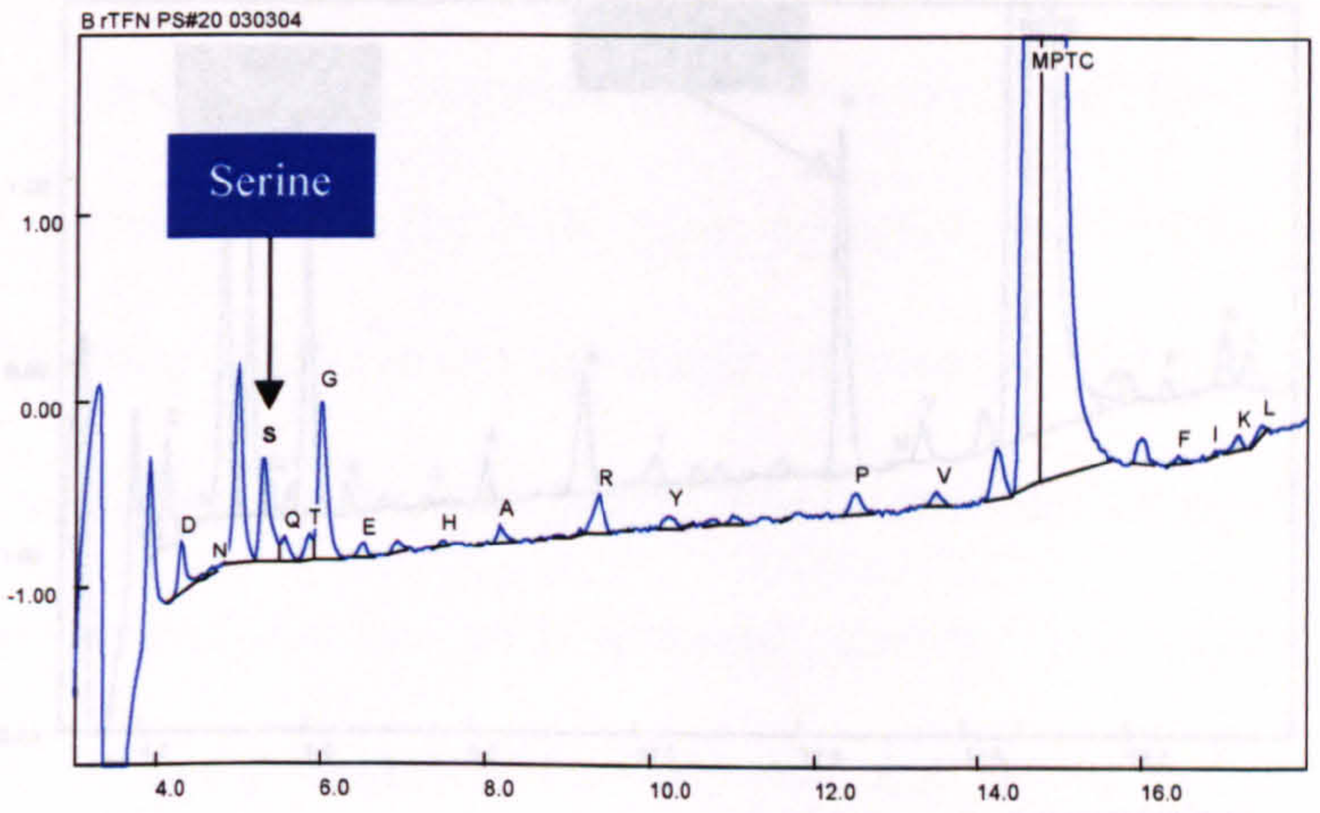


Figure 5.11. FPLC purification of WT rhSTf using a 1ml MonoQ HR 5/5 column and a NaCl gradient of 0 – 250mM (Program 1). a) A trace showing absorbance of eluate at 214nm and percentage of buffer B versus time. Blue line: absorbance and black line: percentage of buffer B. b) Native PAGE showing the protein profile of nine fractions spanning over the major elution peaks.

Cycle 1



Cycle 2



Cycle 3

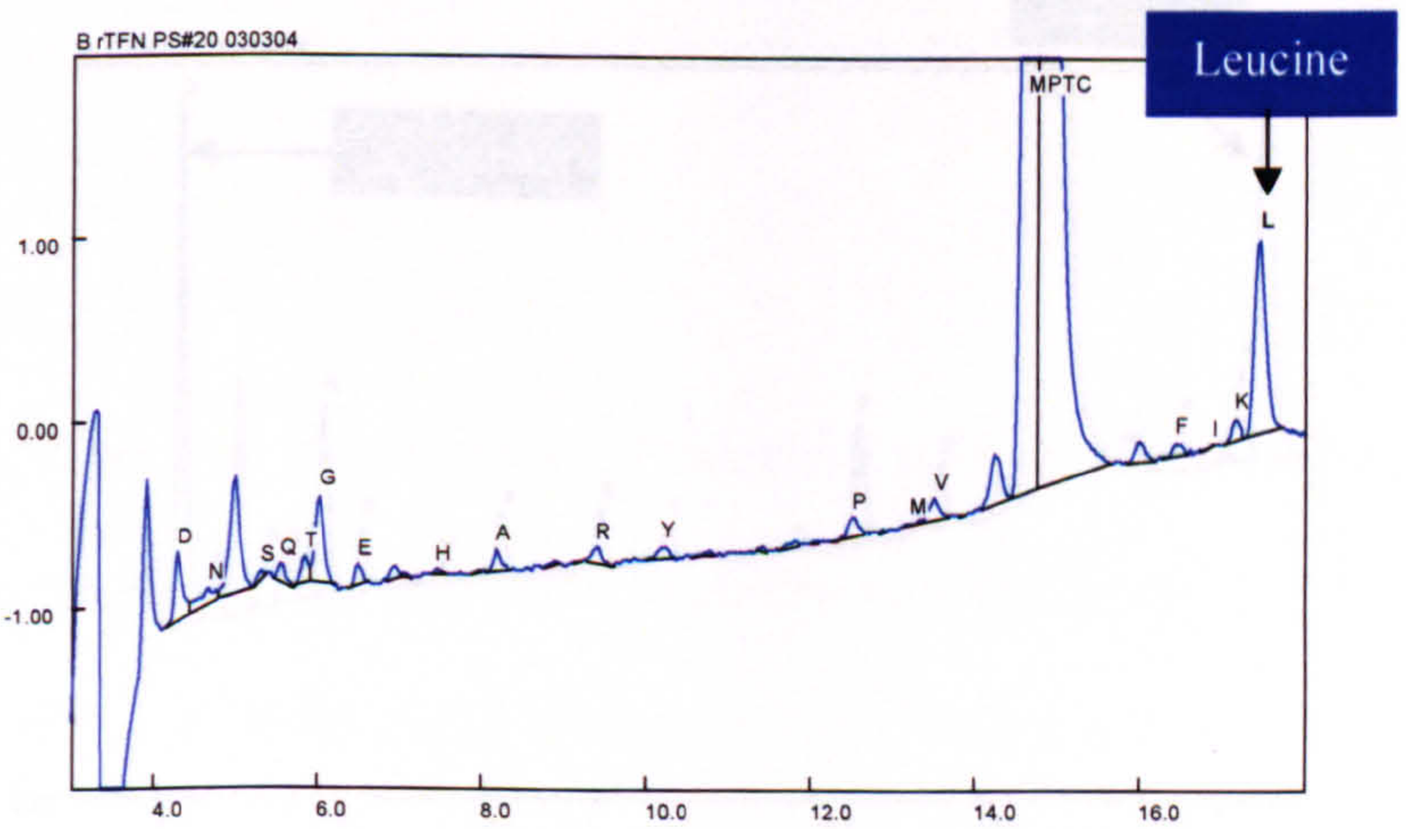
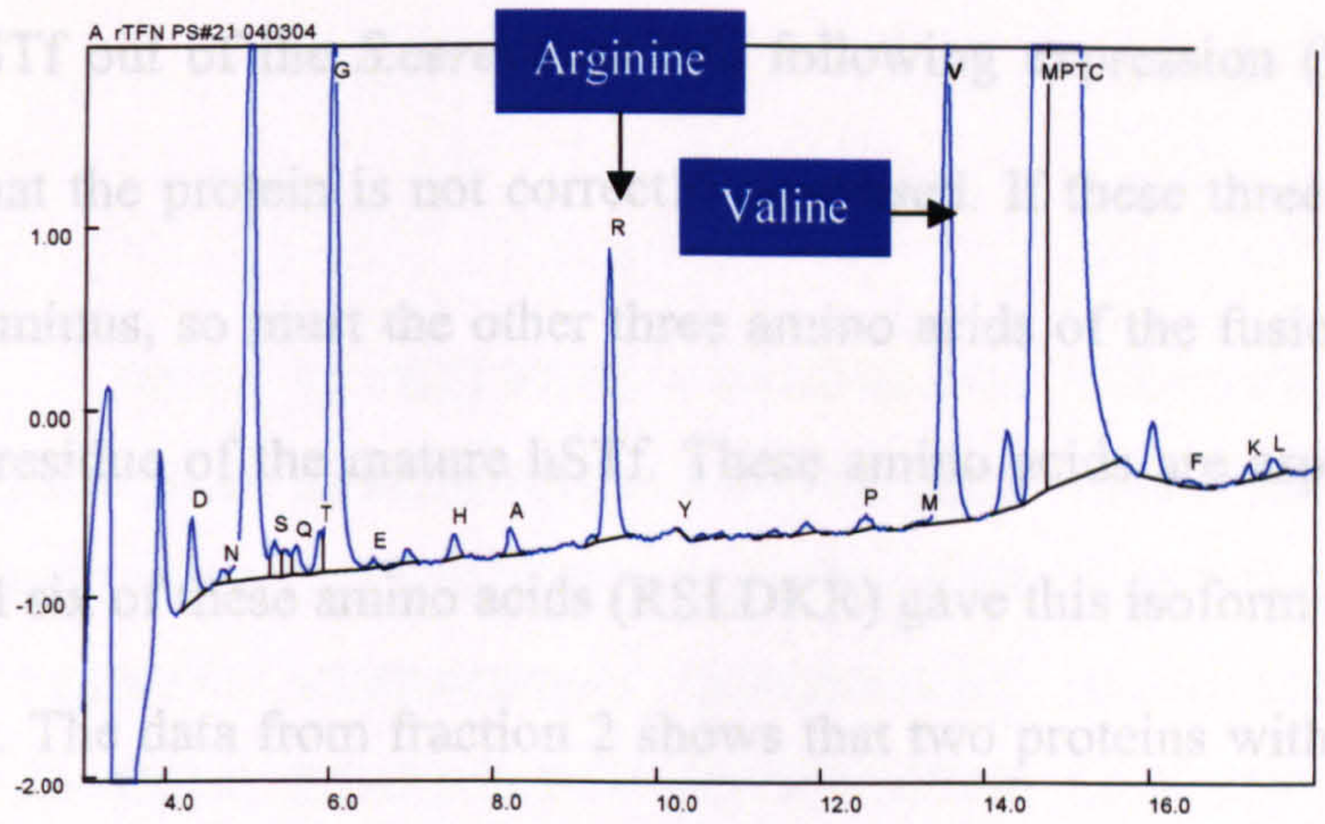
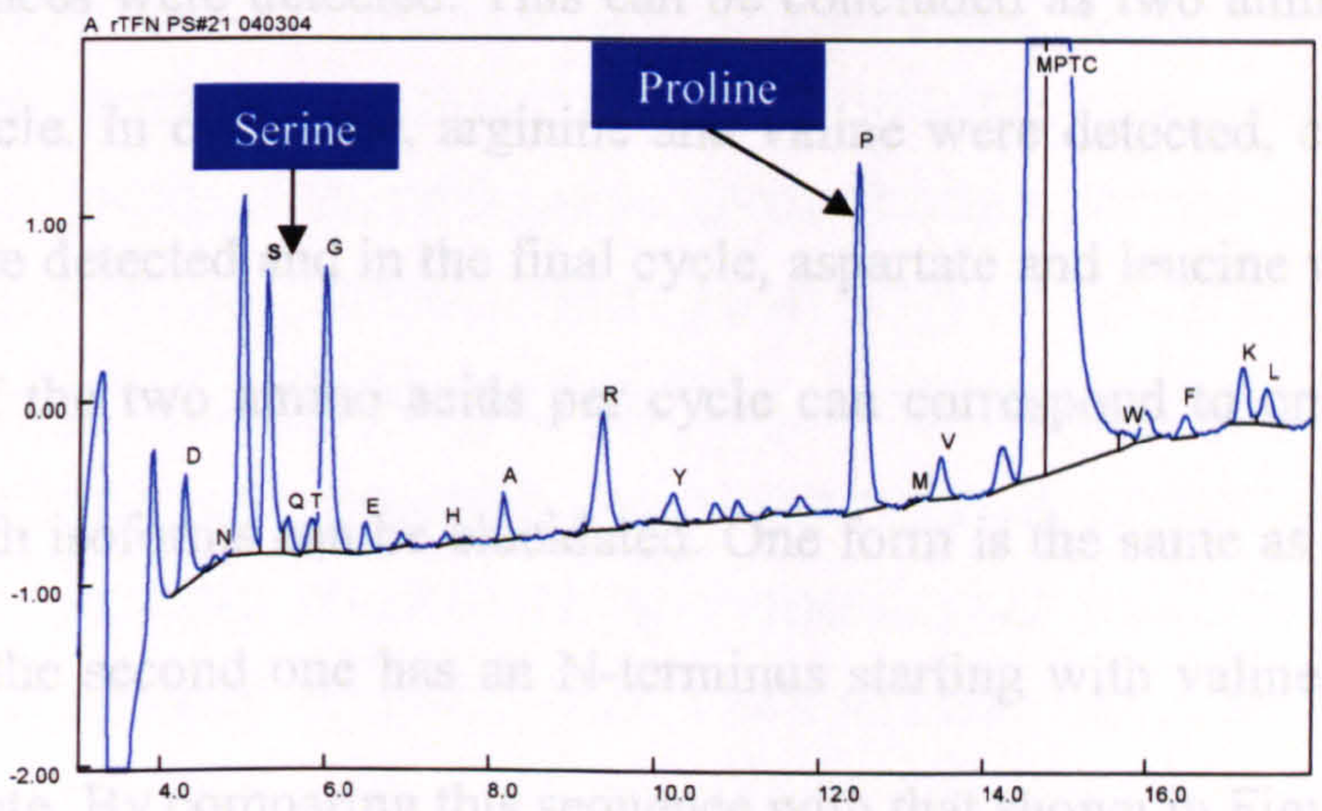


Figure 5.12. N-terminal sequencing of fraction 1 from FPLC MonoQ purification of rhSTf using Program 1.

Cycle 1



Cycle 2



Cycle 3

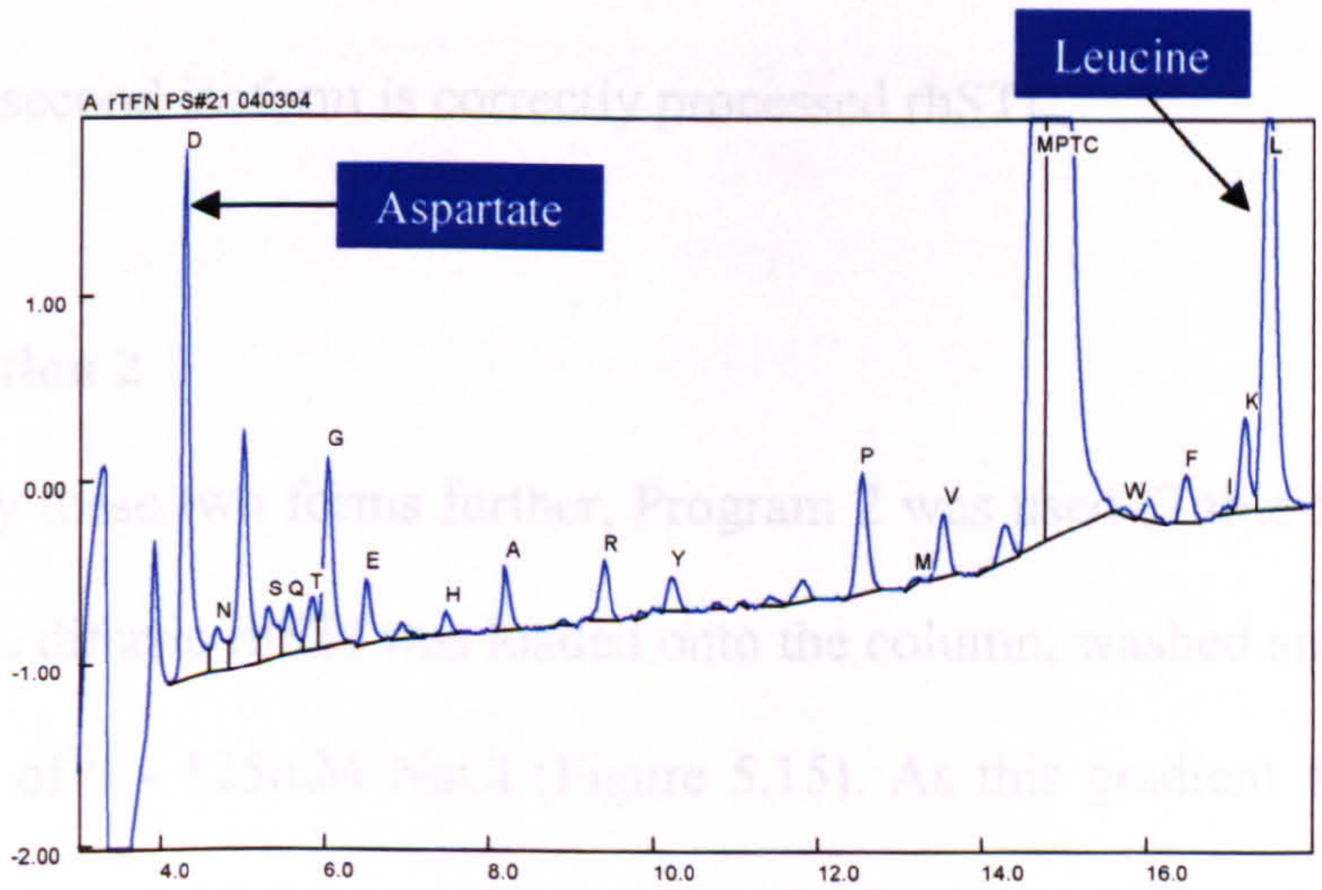


Figure 5.13. N-terminal sequencing of fraction 2 from FPLC MonoQ purification of rhSTf using Program 1.

by serine and then leucine. This sequence is part of the fusion leader used for the secretion of rhSTf out of the *S.cerevisiae* cell following expression (Figure 5.14). This suggests that the protein is not correctly processed. If these three amino acids are on the N-terminus, so must the other three amino acids of the fusion leader that lead up to first residue of the mature hSTf. These amino acids are aspartate, lysine and arginine. All six of these amino acids (RSLDKR) gave this isoform an extra, two positive charges. The data from fraction 2 shows that two proteins with different N-terminus sequences were detected. This can be concluded as two amino acids were released per cycle. In cycle one, arginine and valine were detected, cycle 2, serine and proline were detected and in the final cycle, aspartate and leucine were detected. As only one of the two amino acids per cycle can correspond to one protein, the sequence of both isoforms can be elucidated. One form is the same as the protein in fraction 1 and the second one has an N-terminus starting with valine, then proline and then aspartate. By comparing this sequence with that shown in Figure 5.14 it can be seen that this second isoform is correctly processed rhSTf.

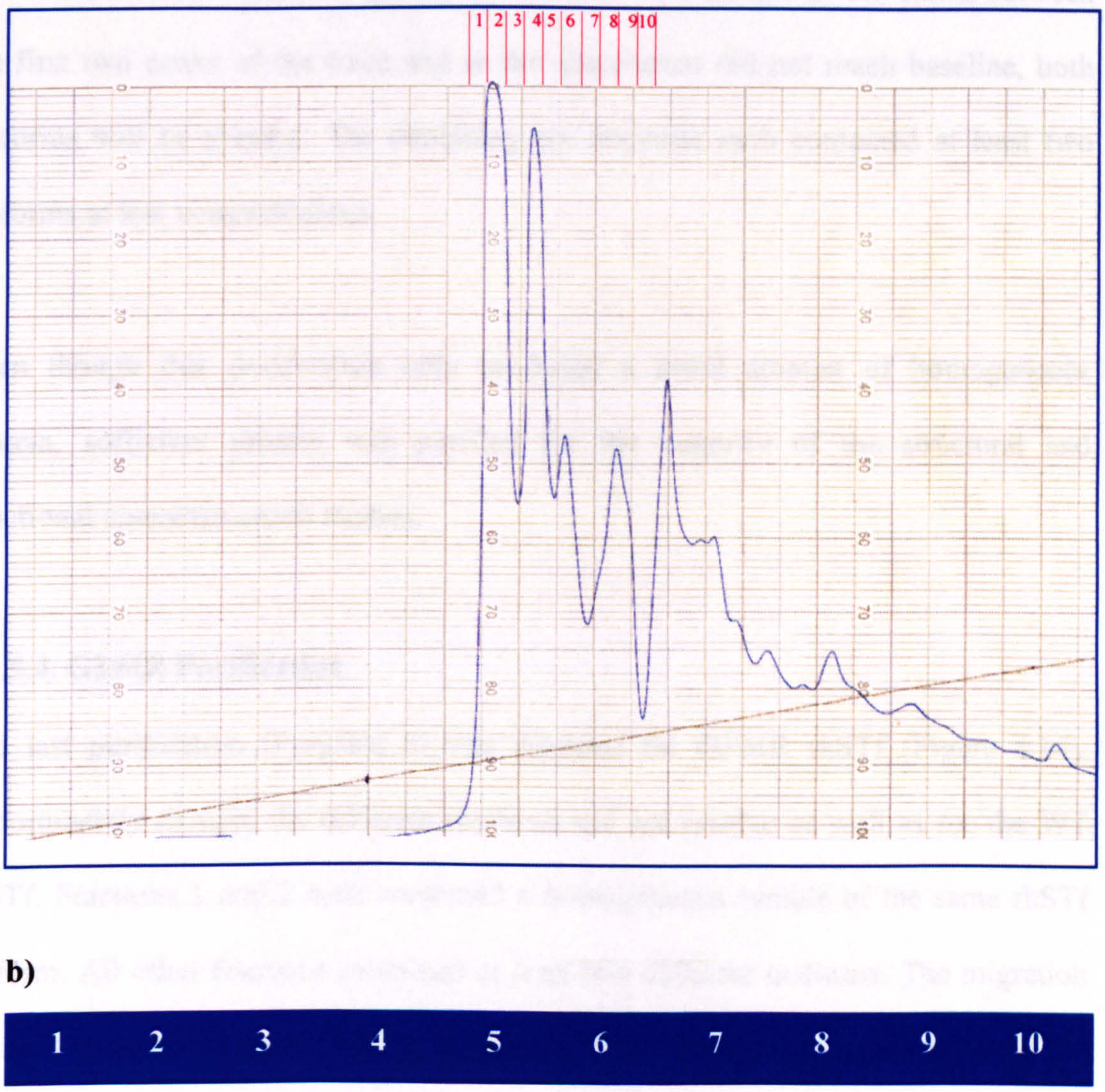
5.1.5.3. Purification 2

In order to purify these two forms further, Program 2 was used (Table 5.1). Another 2mLs of 5mg/mL diferric rhSTf was loaded onto the column, washed and then eluted using a gradient of 0 - 125mM NaCl (Figure 5.15). As this gradient was stretched over the same amount of time as the gradient in Program 1, but the final NaCl concentration was halved, the different isoforms should elute off the column at a greater resolution. This was shown to be successful. Fractions one, two and four contained homogeneous samples of rhSTf. The only difference between these fractions was that number four contained an isoform that migrated further than the

MKWV FIVSILFLFS SAYSRSLDKR VPDKTVRWCA
VSEHEATKCQ SFRDHMKSVI PSDGPSVACV KKASYLDCIR
AIAANEADAV TLDAGLVYDA YLAPNNLKPV VAEFYGSKED
PQTFYYAVAV VKKDSGFQMN QLRGKKSCHT GLGRSAGWNI
PIGLLYCDLP EPRKPLEKAV ANFFSGSCAP CADGTDFPQL
CQLCPGCGCS TLNQYFGYSG AFKCLKDGAG DVAFVKHSTI
FENLANKADR DQYELLCLDN TRKPVDEYKD CHLAQVPSHT
VVARSMGGKE DLIWELLNQA QEHFGKDKSK EFQLFSSPHG
KDLLFKDSAH GFLKVPPRMD AKMYLGYEYV TAIRNLREGT
CPEAPTDECK PVKWCALSHH ERLKCDEWSV NSVGKIECVS
AETTEDCIAK IMNGEADAMS LDGGFVYIAG KCGLVPVLAE
NYQKSDNCED TPEAGYFAVA VVKKSASDLT WDNLKGKKSC
HTAVGRTAGW NIPMGLLYNK INHCRFDEFF SEGCAPGSKK
DSSLCKLCMG SGLNLCEPNN KEGYYGYTGA FRCLVEKGDV
AFVKHQTVPQ NTGGKNPDPW AKNLNEKDYE LLCLDGTRKP
VEEYANCHLA RAPNHAVVTR KDKEACVHKI LRQQQHLFGS
QVTDCSGNFC LFRSETKDLL FRDDTVCLAK LHDRNTYEKY
LGEEYVKAVG NLRKCSTSSL LEACTFRRP

Figure 5.14. Amino acid sequence of WT rhSTf, including the fusion leader. Mature rhSTf sequence: blue and fusion leader: black. Sequence corresponding to the N-terminus of proteins in fraction 1 and 2 are highlighted in yellow.

a)



b)

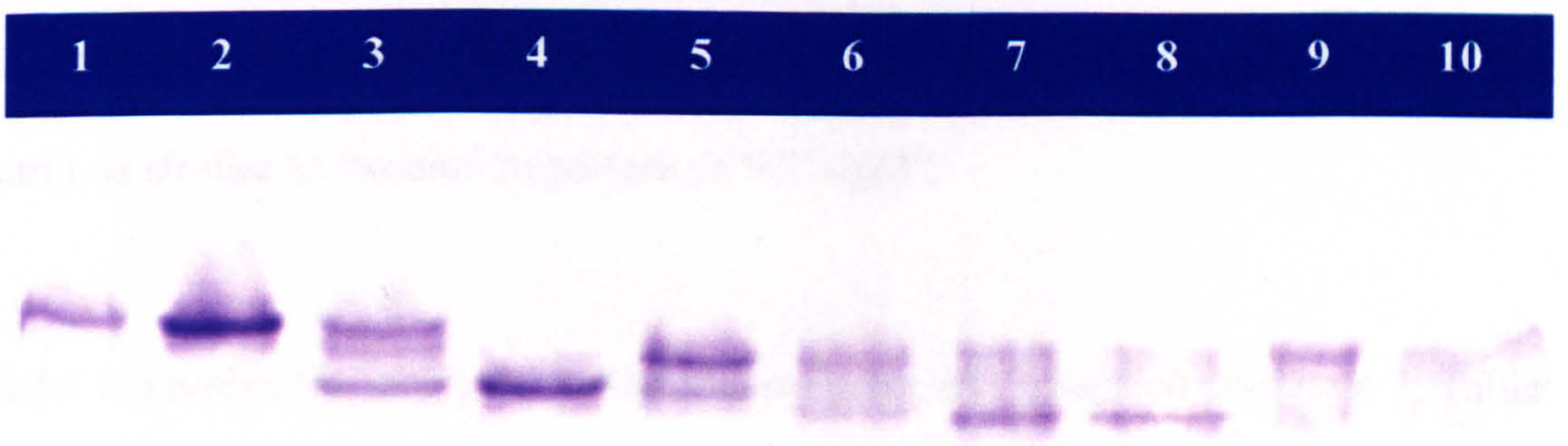


Figure 5.15. FPLC purification of WT rhSTf using a 1ml MonoQ HR 5/5 column and a NaCl gradient of 0 – 125mM (Program 2). a) A trace showing absorbance of eluate at 214nm and percentage of buffer B versus time. Blue line: absorbance and black line: percentage of buffer B. b) Native PAGE showing the protein profile of ten fractions spanning over the major elution peaks.

one seen in the other two fractions. Both of these isoforms are present in fraction 3 at low concentrations. This is expected as fraction 3 is a sample of the eluate between the first two peaks of the trace and as the absorbance did not reach baseline, both isoforms will be present. The remaining six fractions each contained at least two isoforms at low concentrations.

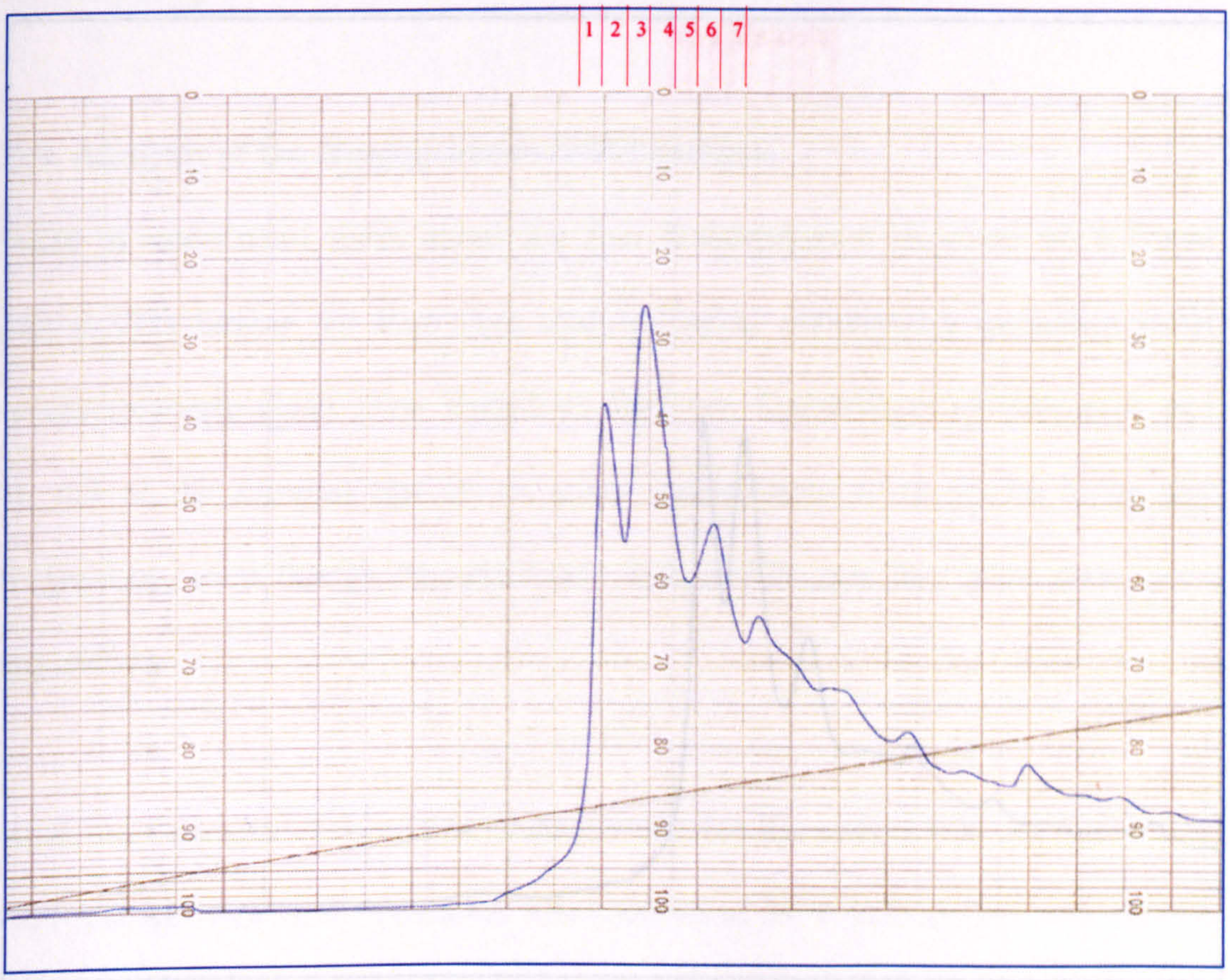
Even though this purification only produced a small amount of homogeneous protein, sufficient protein was purified for the majority of the structural and functional characterisation studies.

5.1.5.4. G394R Purification

The last purification (Program 2) was repeated for G394R rhSTf (Figure 5.16). Unfortunately though, the different isoforms did not resolve as well as for the WT rhSTf. Fractions 1 and 2 both contained a homogeneous sample of the same rhSTf isoform. All other fractions contained at least two different isoforms. The migration of these isoforms on native PAGE, in relation to the time at which they elute off the column, is similar to the elution pattern of WT rhSTf.

In order to resolve the isoforms further, the program was modified (Program 3, Table 5.1) to make the NaCl gradient shallower. By decreasing the NaCl concentration of buffer B from 0.5M to 0.4M, the gradient was reduced from 0 – 100mM over 120min from 0 – 125mM over the same amount of time. Both the elution trace and native PAGE analysis of fractions from this purification are shown in Figure 5.17. Using this new gradient, a second major isoform was purified. This isoform, which eluted off in fraction 4, migrated further on a native PAGE when compared with the

a)



b)

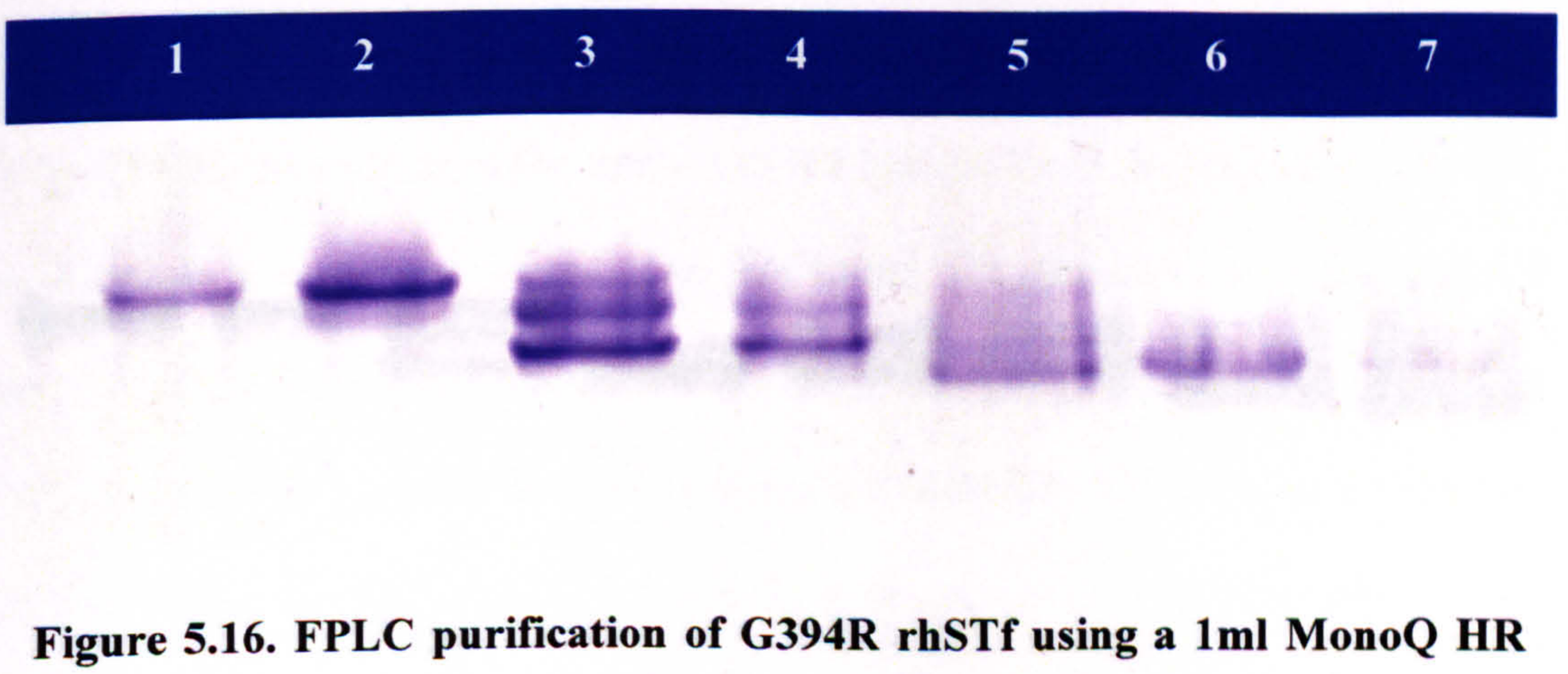
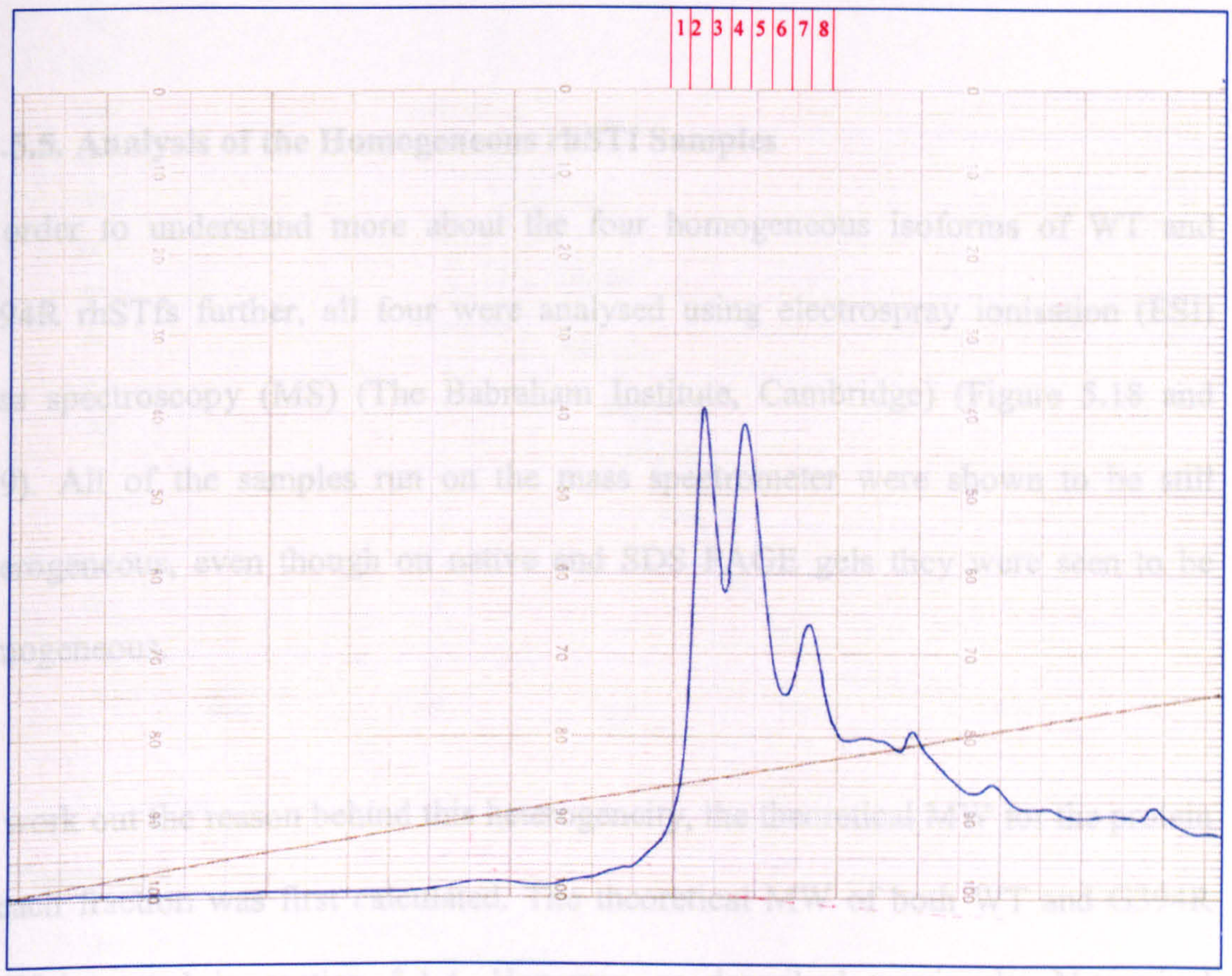


Figure 5.16. FPLC purification of G394R rhSTf using a 1ml MonoQ HR 5/5 column and a NaCl gradient of 0 – 125mM (Program 2). a) A trace showing absorbance of eluate at 214nm and percentage of buffer B versus time. Blue line: absorbance and black line: percentage of buffer B. b) Native PAGE showing the protein profile of seven fractions spanning over the major elution peaks.

a)



b)

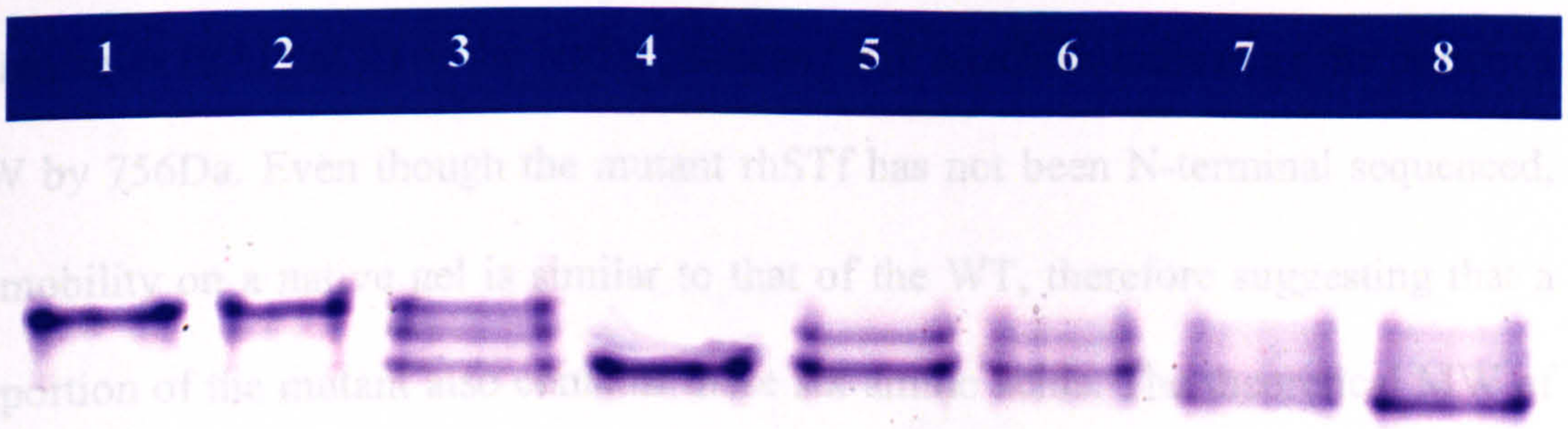


Figure 5.17. FPLC purification of G394R rhSTf using a 1ml MonoQ HR 5/5 column and a NaCl gradient of 0 – 100mM (Program 3). a) A trace showing absorbance of eluate at 214nm and percentage of buffer B versus time. Blue line: absorbance and black line: percentage of buffer B. b) Native PAGE showing the protein profile of eight fractions spanning over the major elution peaks.

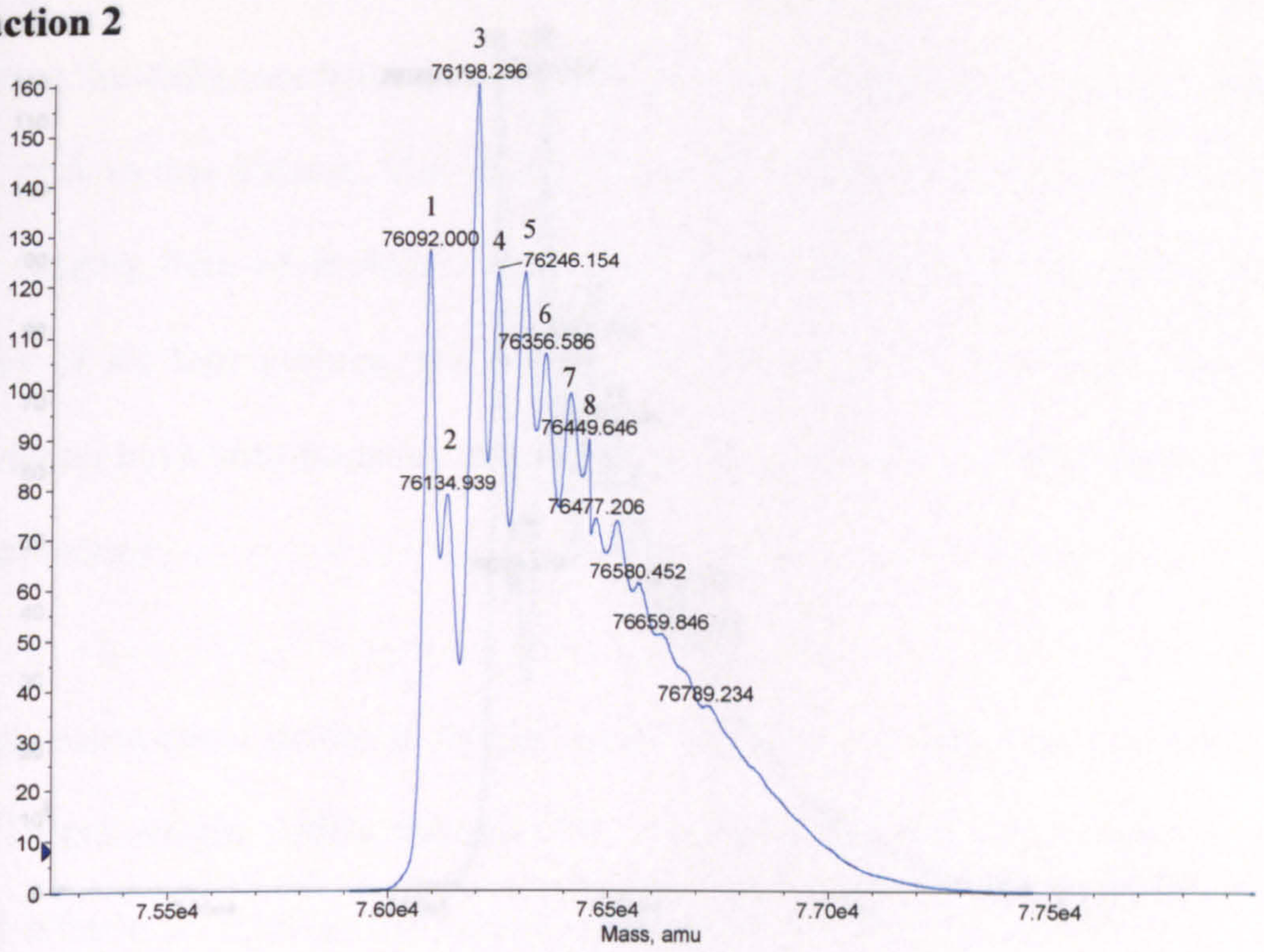
isoform in fractions 1 and 2. Again, this is consistent with the elution pattern of WT rhSTf using program 2, in that two different isoforms were purified.

5.1.5.5. Analysis of the Homogeneous rhSTf Samples

In order to understand more about the four homogeneous isoforms of WT and G394R rhSTfs further, all four were analysed using electrospray ionisation (ESI) mass spectroscopy (MS) (The Babraham Institute, Cambridge) (Figure 5.18 and 5.19). All of the samples run on the mass spectrometer were shown to be still heterogeneous, even though on native and SDS PAGE gels they were seen to be homogeneous.

To work out the reason behind this heterogeneity, the theoretical MW for the protein in each fraction was first calculated. The theoretical MW of both WT and G394R rhSTf is stated in section 5.1.4. However, as described previously, N-terminal sequencing revealed that two forms of WT rhSTf existed. This second form has six amino acids (RSLDKR) of the leader sequence still attached, increasing the protein's MW by 756Da. Even though the mutant rhSTf has not been N-terminal sequenced, its mobility on a native gel is similar to that of the WT, therefore suggesting that a proportion of the mutant also contains these six amino acids. The theoretical MW of the WT and G393R partially processed forms are therefore 75965Da and 76064Da, respectively. For the WT rhSTf, fraction 2 is the partially processed protein and fraction 4 is the fully processed protein. This is the same for the mutant, with fraction 1 being the partially processed protein and fraction 4, the fully processed protein.

Fraction 2



Fraction 4

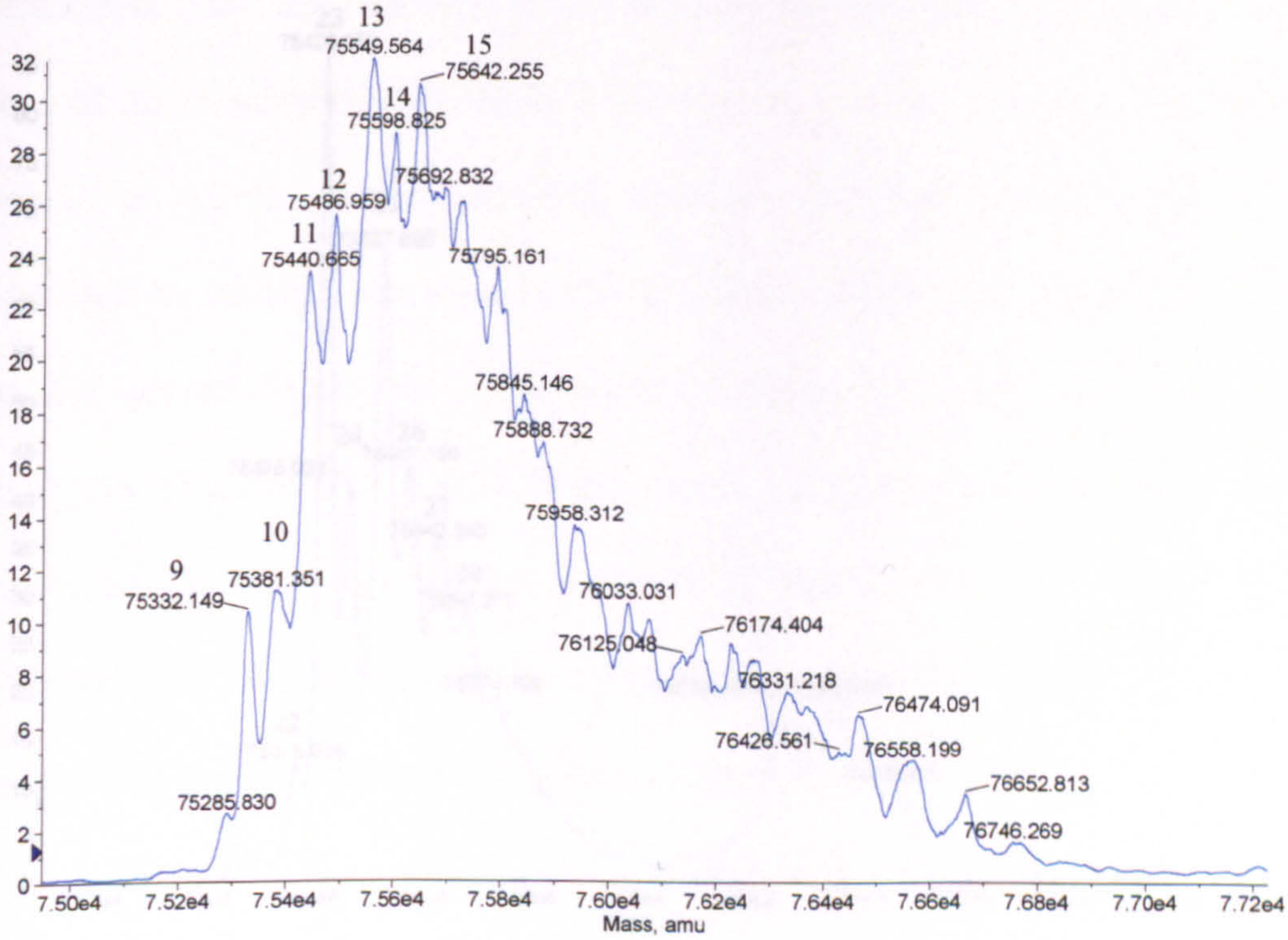
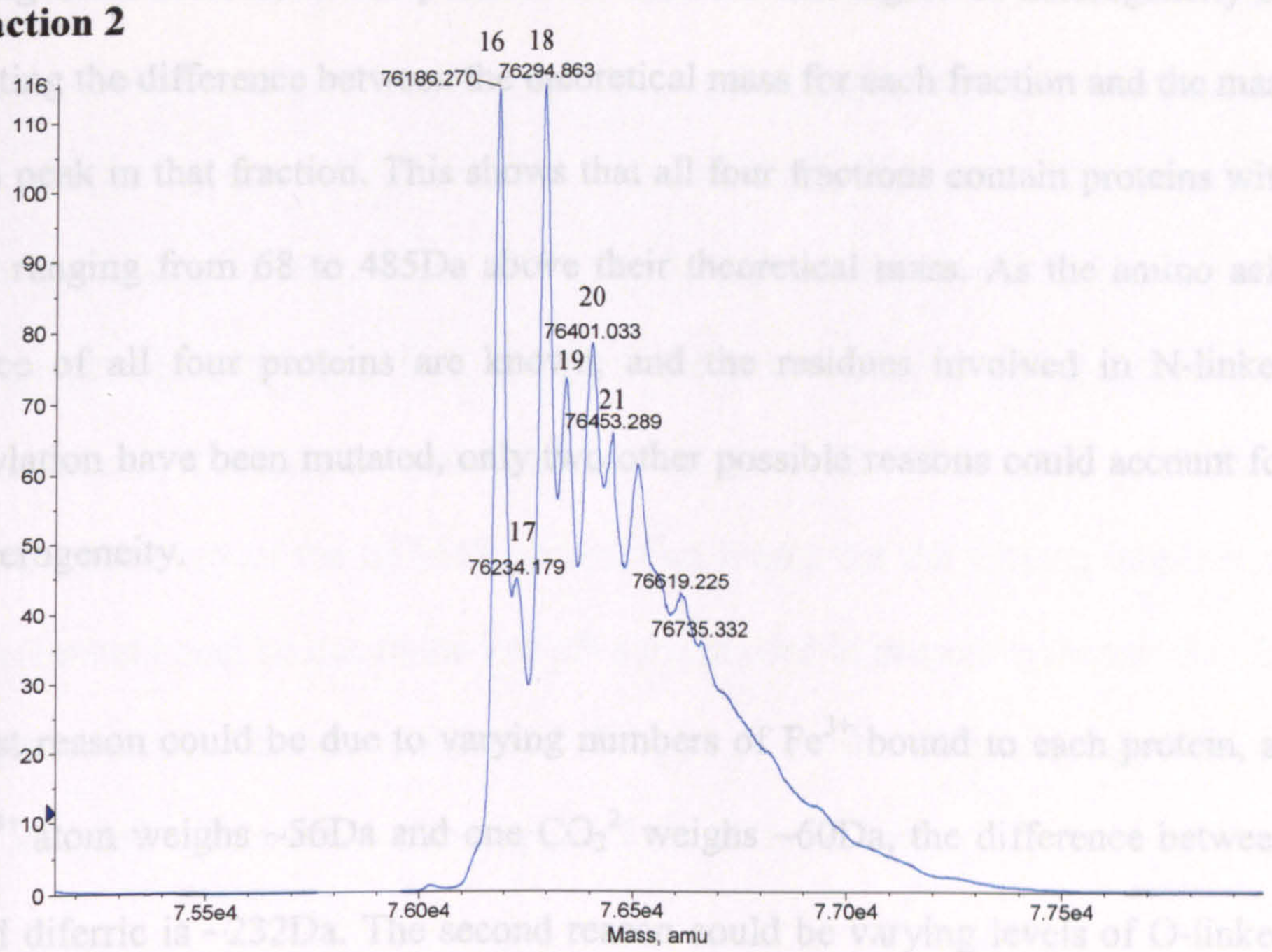


Figure 5.18. Electrospray Ionisation Mass Spectroscopy Analysis of WT rhSTf FPLC fractions using Program 2. Samples were prepared in 20mM ammonium bicarbonate.

Fraction 2



Fraction 4

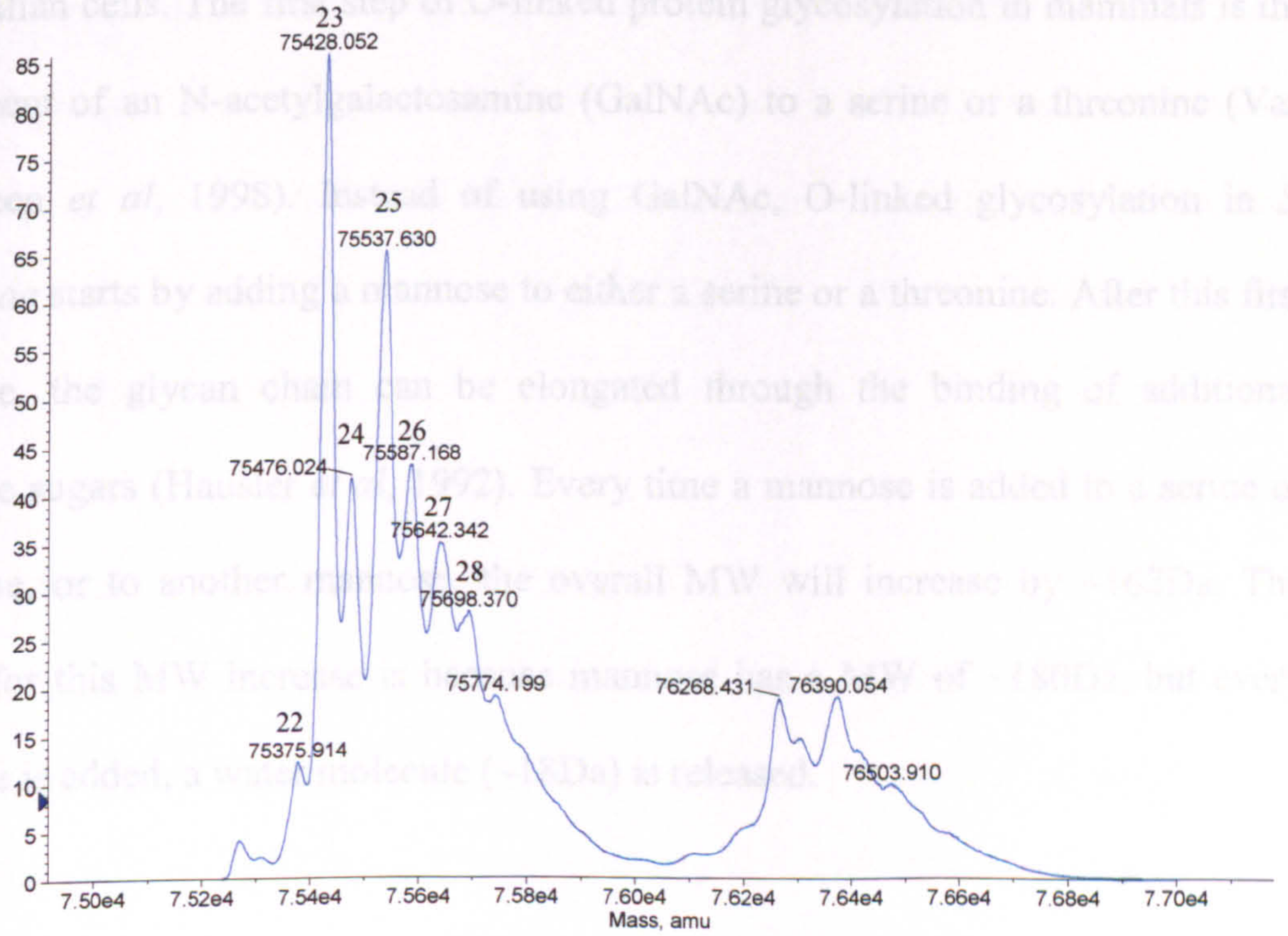


Figure 5.19. Electrospray Ionisation Mass Spectroscopy Analysis of G394R rhSTf FPLC fractions using Program 3. Samples were prepared in 20mM ammonium bicarbonate.

Knowing these masses, it was possible to calculate the degree of heterogeneity by calculating the difference between the theoretical mass for each fraction and the mass of each peak in that fraction. This shows that all four fractions contain proteins with masses ranging from 68 to 485Da above their theoretical mass. As the amino acid sequence of all four proteins are known, and the residues involved in N-linked glycosylation have been mutated, only two other possible reasons could account for this heterogeneity.

The first reason could be due to varying numbers of Fe^{3+} bound to each protein, as one Fe^{3+} atom weighs ~56Da and one CO_3^{2-} weighs ~60Da, the difference between apo and diferric is ~232Da. The second reason could be varying levels of O-linked glycosylation. O-linked glycosylation in *S. cerevisiae* is different to that in mammalian cells. The first step of O-linked protein glycosylation in mammals is the attachment of an N-acetylgalactosamine (GalNAc) to a serine or a threonine (Van den Steen *et al*, 1998). Instead of using GalNAc, O-linked glycosylation in *S. cerevisiae* starts by adding a mannose to either a serine or a threonine. After this first mannose, the glycan chain can be elongated through the binding of additional mannose sugars (Hausler *et al*, 1992). Every time a mannose is added to a serine or threonine, or to another mannose, the overall MW will increase by ~162Da. The reason for this MW increase is because mannose has a MW of ~180Da, but every time one is added, a water molecule (~18Da) is released.

By working out the differences between the mass of the first peak and the theoretical mass (calculated using Compute pI/MW tool, us.expasy.org/tools), and then between each peak from then on, the number of O-linked mannoses attached to each protein

and the iron status of that protein can be determined (Table 5.2). Between six to eight peaks per fraction are labelled. The smallest peak in each fraction relates to the protein with no mannoses attached, but with one Fe^{3+} ion bound. This protein is ~116Da larger than its theoretical MW. The largest peak in each fraction usually represents protein with more than one mannose attached and varying numbers of Fe^{3+} ions bound. It should be noted that on average a 6Da (0.007% of total mass) error in the accuracy of the ESI-MS occurs. One reason for the varying numbers of iron ions bound could be due to the low pH solvent used in preparing the samples for mass spectroscopy. Therefore, some of the protein in the diferric preparation of protein analysed will have lost either one, or both of its Fe^{3+} ions.

By comparing the MW of proteins with the same number of mannoses and Fe^{3+} ions attached, a mass difference of ~756Da can be observed between the two fractions of the same protein. The difference between peak 2 and 10 is 754Da and the difference between peak 17 and 24 is 758Da. These differences are consistent with the presence of the partially processed leader sequence RSLDKR in fraction 2 of the WT and fraction 1 of the mutant.

5.1.5.6. Scaled-up MonoQ FPLC Purification of rhSTf

In order to purify a larger quantity of the four isoforms described above, the 1mL monoQ column was replaced by a 10mL monoQ column. A number of modifications were made to these purifications. The flow rate was increased from 0.5ml/min to 2ml/min, and instead of loading 5mg of diferric protein, 20mg was loaded. All other parameters were kept the same.

Protein	Theoretical size (Da)	Fraction	Peak No.	Size (Da)	Iron status *+116 Da per Fe^{3+} + CO_3^{2-}*	Number of Mannoses *+162 Da per mannose*	Error (Da)
WT	75965	2	1	76092	1	0	+11
			2	76135	0	1	+8
			3	76198	2	0	+1
			4	76246	1	1	+3
			5	-	-	-	-
			6	76357	2	1	+2
			7	-	-	-	-
			8	76450	0	3	-1
	75209	4	9	75332	1	0	+7
			10	75381	0	1	+10
			11	75441	2	0	0
			12	75487	1	1	0
			13	75550	0	2	+17
			14	75599	2	1	-4
			15	75642	1	2	-7
G394R	76064	2	16	76186	1	0	+6
			17	76234	0	1	+8
			18	76295	2	0	-1
			19	-	-	-	-
			20	76401	0	2	+13
			21	76453	2	1	-5
	75308	4	22	75376	-	-	-
			23	75428	1	0	-4
			24	75476	0	1	+6
			25	75538	2	0	-2
			26	75567	1	1	-19
			27	75642	0	2	+10
			28	75698	2	1	-4
					Average error (+/-Da)		*6*

Table 5.2. Theoretical number of O-linked mannoses attached to each protein visualised using ESI mass spectroscopy (figure 5.18 and 5.19) and their theoretical iron status. This was calculated by firstly working out the difference between the peak mass and its theoretical mass. The number of mannoses (162Da), and or Fe^{3+} + CO_3^{2-} (116Da) that could make up this mass difference was calculated empirically.

5.1.6. Native G394R hSTf Purification

5.1.6.1. Immunoaffinity Chromatography

50mL of serum from the individual who is heterozygous for the G394R hSTf mutation was centrifuged at 10,000g for 10min and then filtered through a 0.2μ filter. The serum was then diluted 1:2 with immunoaffinity chromatography equilibration buffer before being run on the anti-hSTf affinity column described in section 5.1.3.2 As with previous purifications using this column, samples from the load, the load flow-through, the wash flow-through and the elution were analysed using SDS-PAGE (Figure 5.20). This Figure shows that a pure preparation of hSTf was purified from the serum. As before, eluted protein was desalted using a G-25 column and freeze-dried immediately following purification.

5.1.6.2. Fast Performance Liquid Chromatography Purification

In order to separate the mutant transferrin from the WT transferrin from this affinity purified preparation, 5mg of it was run on a 1mL monoQ column (section 5.1.5.1). Using the same protocol described in section 2.3.5, this preparation was made diferric and then filtered through a 0.2μ filter before being injected onto the 2mL loading-loop. Program 4 (Table 5.1) was then used to load and purify this sample using a gradient of 0 - 200mM NaCl (Figure 5.21a). Samples from nine fractions spanning over the elution peak were collected and visualised using native PAGE (Figure 5.21b). The first four fractions visualised on a native PAGE contained a single isoform that migrated faster than the isoform in the last four fractions. Fraction 5, contained a mixture of the two isoforms. These two isoforms correspond to the WT (fractions 1-4) and mutant (fractions 6-9) transferrin. The fact that the mutant runs slower on a native PAGE is discussed in section 5.1.3.3. However, the fact that

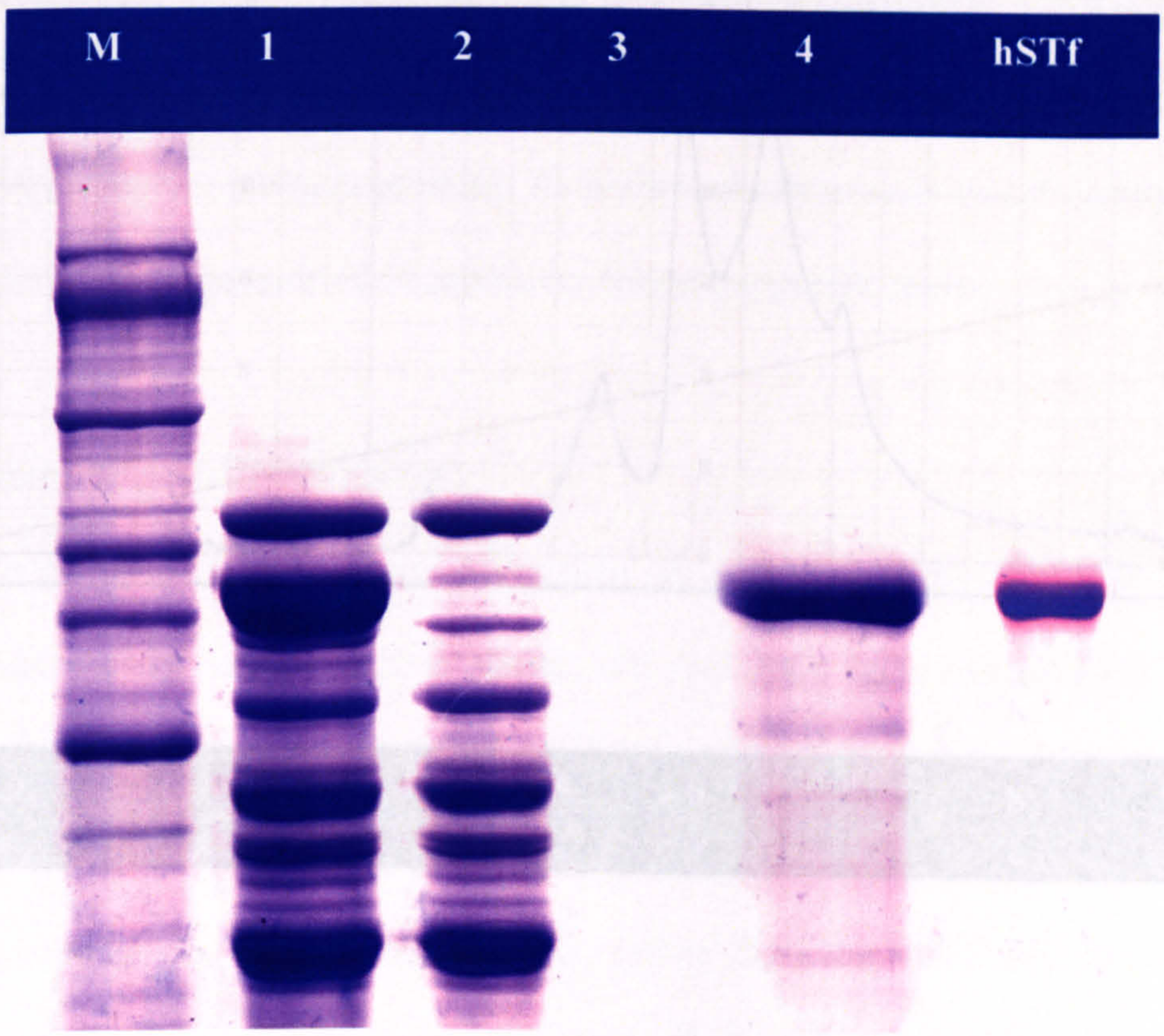
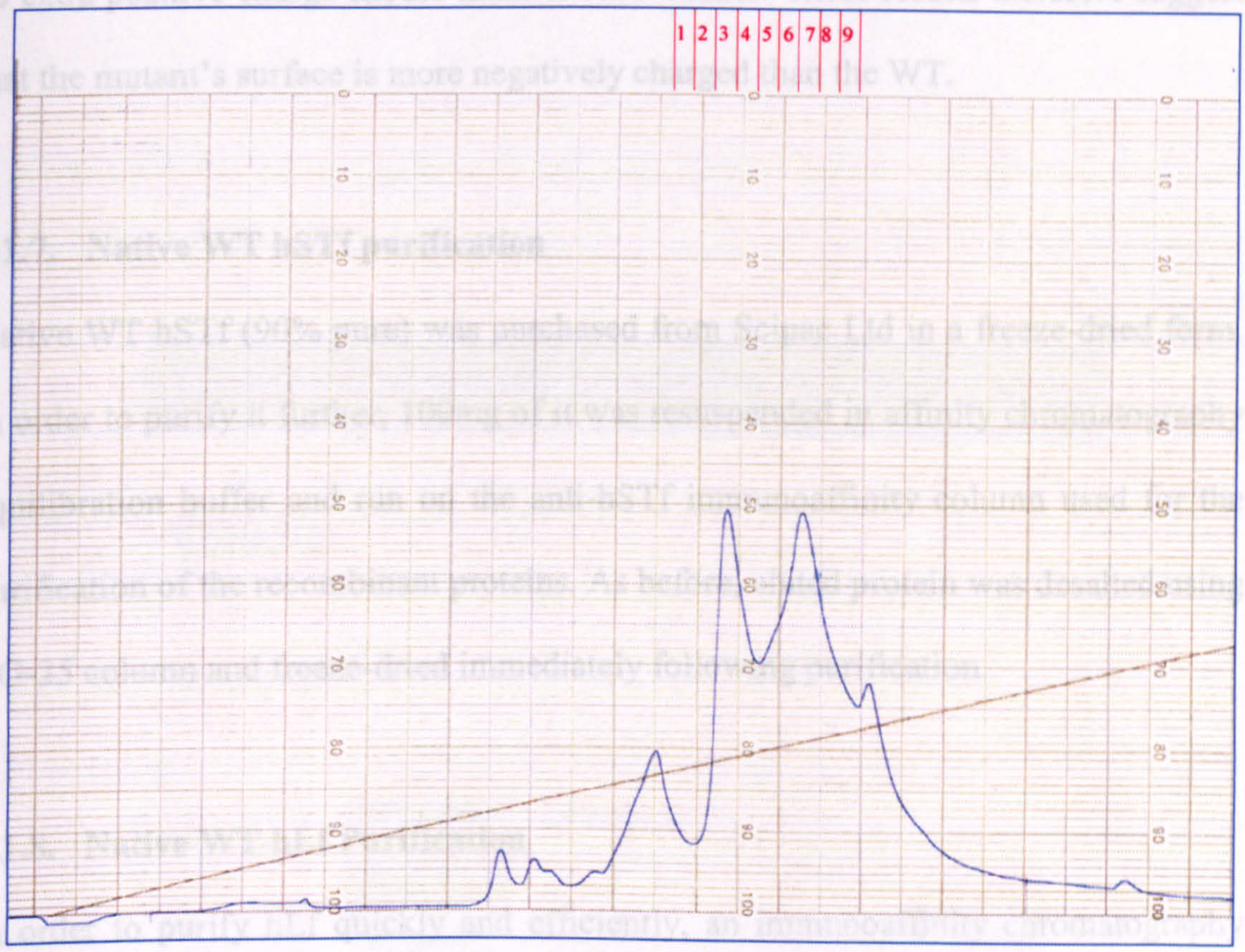


Figure 5.20. Reducing SDS PAGE of samples collected during native G394R hSTf purification using an anti-hSTf affinity column. M: HMW-SDS marker, 1: load, 2: load flow-through, 3: wash, 4: Eluate and hSTf: 1μg of native hSTf (Scipac Ltd)

a)



b)

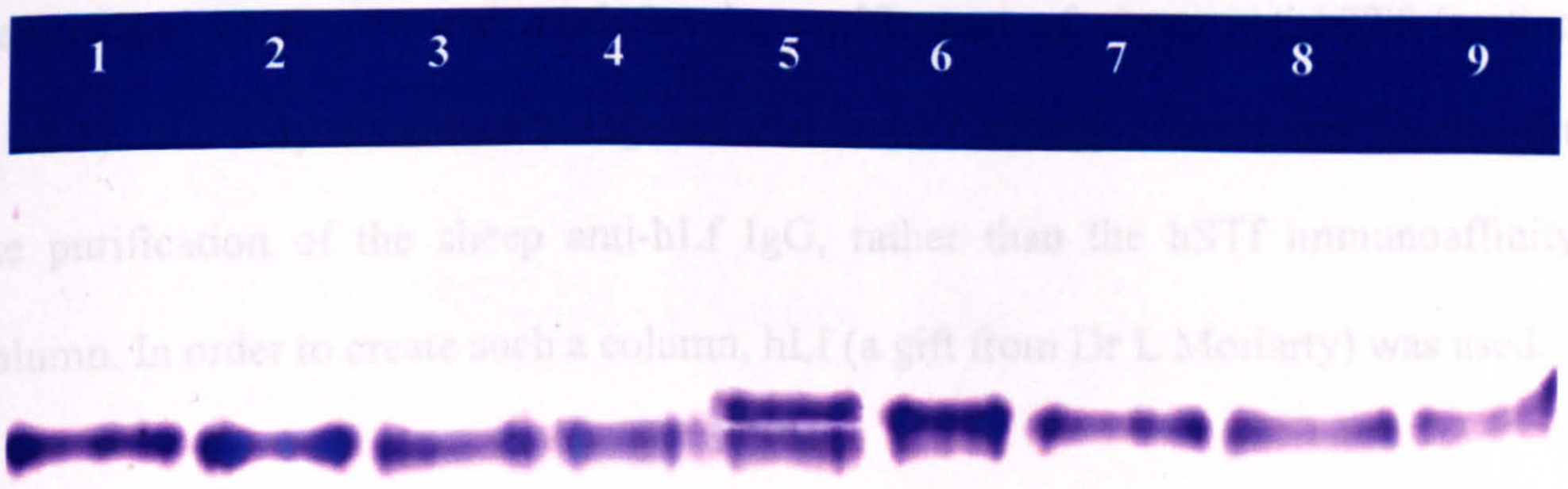


Figure 5.21. FPLC purification of native G394R hSTf using a 1ml MonoQ HR 5/5 column and a NaCl gradient of 0 – 200mM (Program 4). a) A trace showing absorbance of eluate at 214nm and percentage of buffer B versus time. Blue line: absorbance and black line: percentage of buffer B. b) Native PAGE showing the protein profile of nine fractions spanning over the major elution peaks.

the mutant elutes of the monoQ column later than the WT protein is a mystery since its extra positive charge should make it bind weaker. These results therefore suggest that the mutant's surface is more negatively charged than the WT.

5.1.7. Native WT hSTf purification

Native WT hSTf (90% pure) was purchased from Scipac Ltd in a freeze-dried form. In order to purify it further, 100mg of it was resuspended in affinity chromatography equilibration buffer and run on the anti-hSTf immunoaffinity column used for the purification of the recombinant proteins. As before, eluted protein was desalted using a G-25 column and freeze-dried immediately following purification.

5.1.8. Native WT hLf Purification

In order to purify hLf quickly and efficiently, an immunoaffinity chromatography column was created using purified sheep anti-hLf IgG. IgG was purified using exactly the same principal used for the purification of sheep anti-hSTf (section 5.1.3.1). The only difference being was that an hLf affinity column was created for the purification of the sheep anti-hLf IgG, rather than the hSTf immunoaffinity column. In order to create such a column, hLf (a gift from Dr L Moriarty) was used.

Following the creation of the anti-hLf affinity column, 200mL of unpasteurised human milk (donated by King's College Hospital Neonatal Unit, London) was centrifuged at 5000g for 40mins. The resulting cream layer was removed by filtration through glass wool. Using the protocol described in section 2.3.2, hLf was then purified from the resulting supernatant. The purity of hLf in the eluate was determined by SDS PAGE (Figure 5.22) to be highly pure.

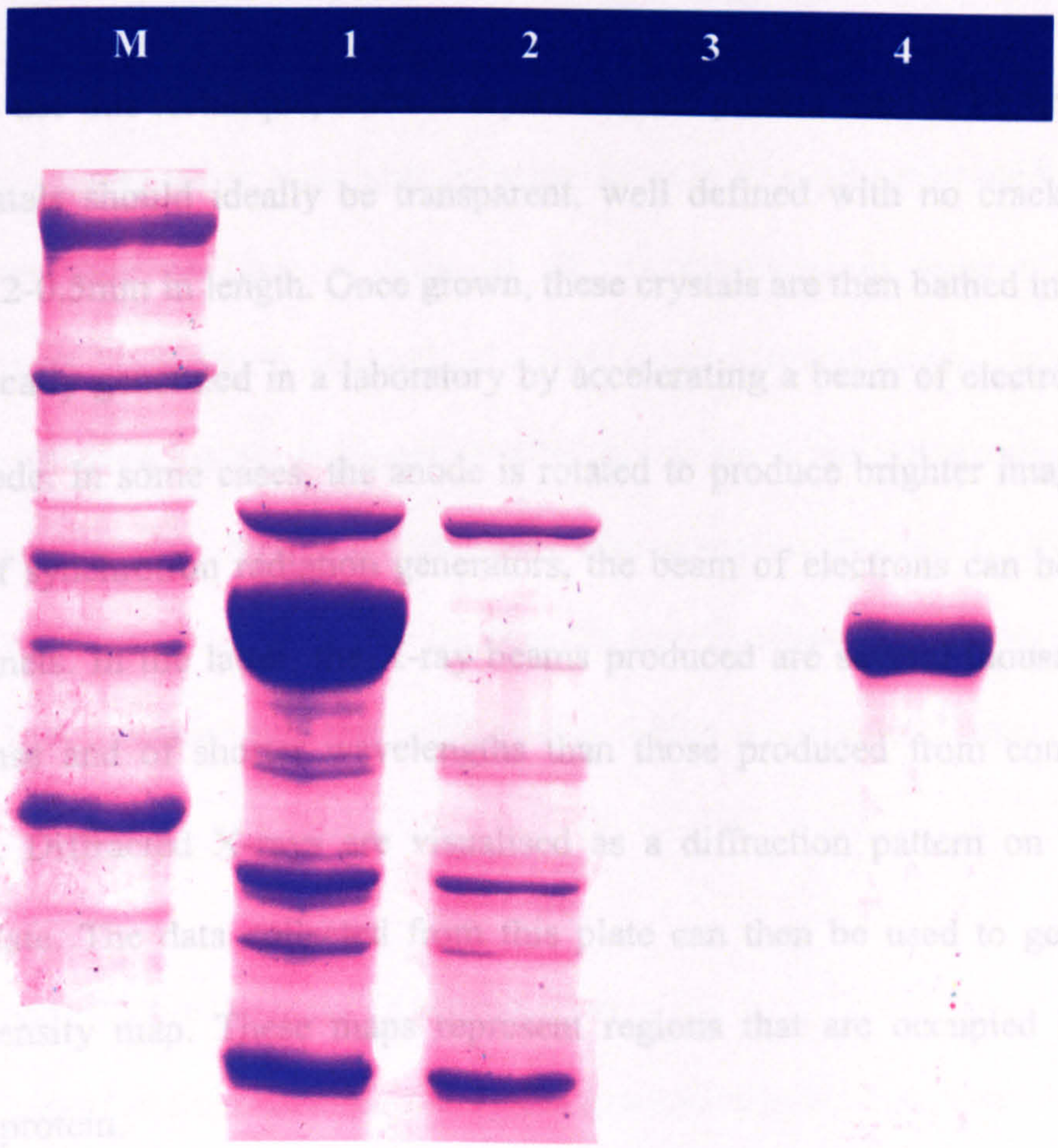


Figure 5.22. Reducing SDS-PAGE of samples collecting during hLf purifications using an anti-hLf affinity column. M: HMW-SDS marker, 1: Load, 2: Load flow-through, 3: Wash and 4: Eluate.

5.2. X-ray Crystallography of hLf and hSTf

X-ray crystallography is perhaps the most powerful technique for determining protein structure and has been used extensively for decades. This technique relies on the interference, or diffraction, caused by X-rays being scattered by electrons surrounding atoms aligned within the protein crystals.

In order to use this technique, suitable crystals of the protein need to be first grown. These crystals should ideally be transparent, well defined with no cracks and be between 0.2-0.5mm in length. Once grown, these crystals are then bathed in an X-ray beam, typically generated in a laboratory by accelerating a beam of electrons into a cooled anode. In some cases, the anode is rotated to produce brighter images, or in the case of synchrotron radiation generators, the beam of electrons can be focused using magnets. In the latter, the X-ray beams produced are several thousand times more intense and of shorter wavelengths than those produced from conventional generators. Diffracted X-rays are visualised as a diffraction pattern on an X-ray detector plate. The data collected from this plate can then be used to generate an electron density map. These maps represent regions that are occupied by atoms within the protein.

As discussed in section 1.3.2 the tertiary structure of many of the transferrins have been elucidated using X-ray crystallography. However, to date, no high-resolution image of hSTf has been obtained. The reason for the failure to stabilise hSTf is still a bit of a mystery, but one possible reason could be the heterogeneity in the N-linked glycosylation found on the protein. Therefore, the initial aim of this aspect of the study is to explore conditions for the crystallisation of recombinant hSTf.

The second aim of this study is to explore conditions for the crystallisation of hLf bound to indium(III). Even though hLf has been crystallised previously in both its apo- (Jameson *el al*, 1998) and diferric form (Haridas *et al*, 1995) there are still many questions yet to be answered which could help us understand the mechanics of all proteins in the Transferrin family.

One of these questions is how both domains, in each lobe, close to bind the metal ion, whether they do open or close at all. As discussed in section 1.3.1, the Transferrins are known to be able to bind a wide range of metal ions of varying sizes. One of the smallest known is aluminium(III) which has a molecular radii of 57pm and the largest is cerium(III), a lanthanide with a molecular radii of 99pm. This shows the residues involved in iron binding are able to adapt to different size metal ions. How they are able to do this is a question, which we attempt to answer in these studies.

5.2.1. Introduction to Protein Crystallisation

Crystallisation of proteins is achieved by the gradual removal of solvent around the protein leading to state known as supersaturation. This state can be split up into three phases. The first phase is known as the Metastable phase and is where crystals may not form for a long time, but the conditions are suitable to sustain growth. Phase two, or the Nucleation phase, is where protein crystals nucleate and grow and finally, the Precipitation state. This is where proteins will precipitate out of solution. The rate at which this supersaturation is attained is dependent on many variables. These include

protein concentration, pH, temperature, as well as the nature and concentration of precipitant present.

Two main methods are commonly used for the crystallisation of serum transferrins and lactoferrins: the dialysis and the vapour diffusion method. The latter is the most common, but both of these methods work using the same principal. With the dialysis method, the protein is dissolved in buffer and precipitant and then dialysed against a higher concentration of precipitant. This gradually removes the water around the protein causing supersaturation. The rate of attainment of supersaturation can be controlled by varying the difference in concentration of precipitant on either side of the dialysis membrane. With the vapour diffusion method (Hanging Drop), a droplet of protein solution containing precipitant is equilibrated with a reservoir of precipitant containing a higher concentration of precipitant. Equilibration proceeds in this case by the diffusion of solvent from the drop to the reservoir, thus achieving supersaturation in the drop. This occurs until the vapour pressures in the drop and the reservoir are equal.

During this study, only vapour diffusion was explored for the crystallisation of hLf. However, a third technique was used in attempts to crystallise hSTf. This third technique, known as Microbatch, involves mixing the protein sample with the crystallisation agent and then placing a drop of this under oil. Unlike vapour diffusion and the dialysis method, the concentration of the ingredients in the crystallisation agent is such that supersaturation is achieved immediately upon mixing, thus the composition and the volume of a trial remain constant and crystals will only form if the precise conditions have been correctly chosen. The benefits of

this technique are that very small amounts of sample are required and crystallisation conditions are less prone to experimental error.

5.2.2. Crystallisation of Diferric Recombinant hSTf

As discussed previously, to date, no high-resolution image of diferric-hSTf has been obtained. Therefore, in this section, conditions for the crystallisation of recombinant hSTf were explored in the hope of obtaining a high-resolution image of this protein. The growth of hSTf crystals and subsequent X-ray diffraction analysis in this study was carried out in collaboration with Dr Claire Naylor and Dr Bernard O'Hara, Department of Crystallography, Birkbeck College, London.

Recombinant WT hSTf, purified and fractionated in section 5.1.5.6, was iron loaded, as described in section 2.3.5, and then desalted and freeze-dried as described in section 2.3.6. Prior to use, proteins were then resuspended into the relevant buffer. All hSTf crystallisation assays were carried out using the Microbatch assay described in section 2.4.11.

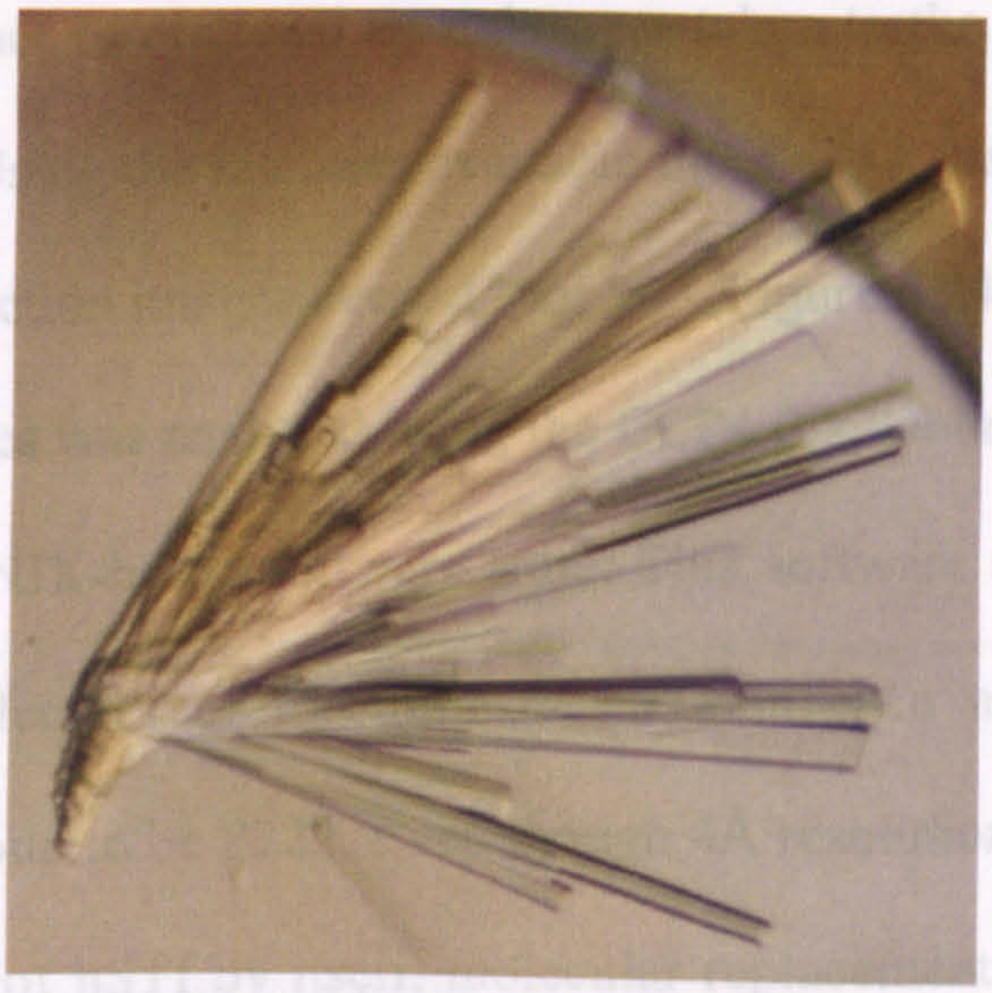
The first screen used the Hampton Research Crystal Screen 1 (Appendix III), with a protein concentration of 3mg/mL. It should be noted that since 1μL of protein is mixed with 1μl of crystallisation agent, the final concentration of protein is around 1.5mg/ml. These crystallisation screens were incubated at 293K for 2 weeks, but no crystals grew.

In order to try and increase the chance of crystal growth, this first screen was repeated, but with the protein concentration at 5.5mg/ml. Also, using the same

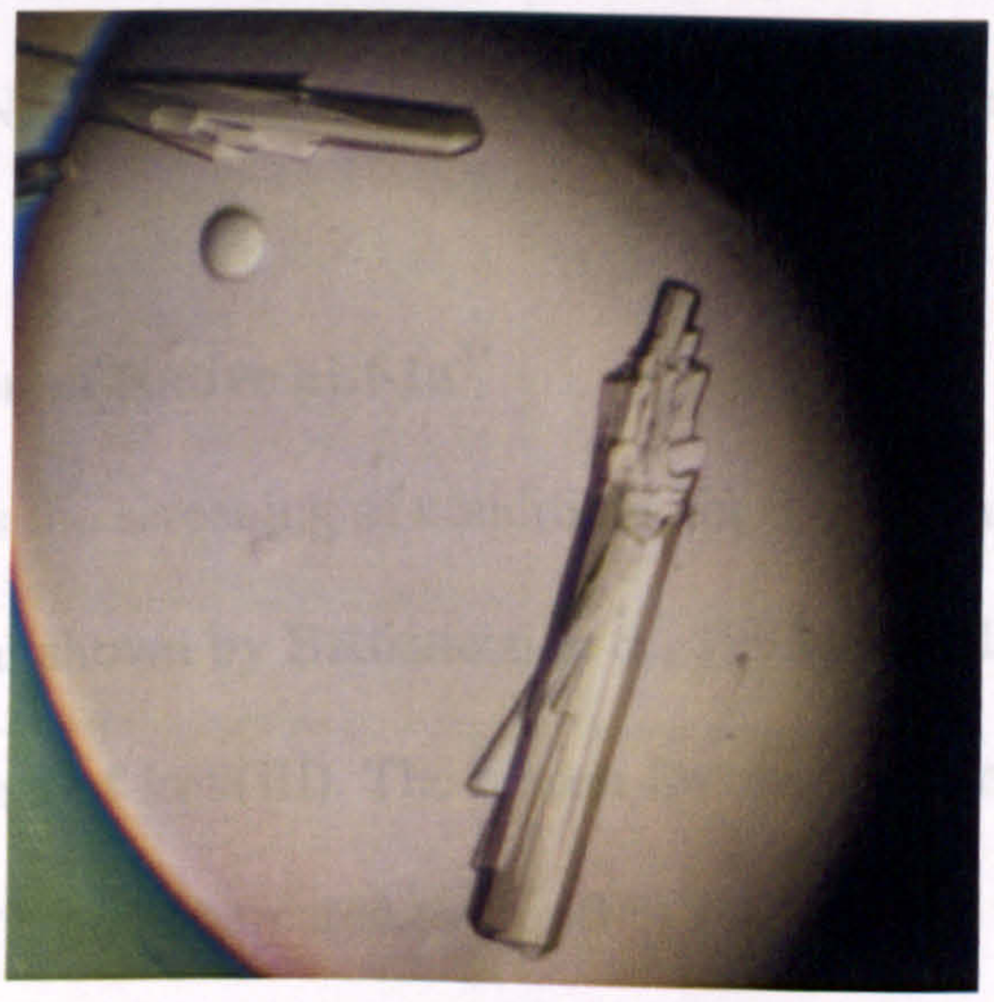
protein concentration, a second screen (Hampton Research Crystal Screen 2, Appendix III) was also explored. From these two screens, crystals grew in two of the crystallisation conditions, following incubation at 293K for approximately 2 weeks (Figure 5.23). These two conditions are screen 1, buffer 40: 0.05M tri-sodium citrate dihydrate pH 5.6, 20 %v/v iso-propanol and 20% (w/v) PEG 4000 and screen 2, buffer 39: 0.2M magnesium chloride hexahydrate, 0.1M tris (pH 8.5), 3.4M 1,6 hexanediol. It should be noted that as hSTf has a pI of around 5.2 (tasialotransferrin) to 5.7 (disialotransferrin), with tetrasialotransferrin and trisialotransferrin residing in between (Chang *et al*, 2005), the ideal pH for crystallisation should be around pH 5.5 Therefore, it is not surprising that crystals grew in buffer 40, screen 1, but it is surprising that they grew in buffer 39, screen 2. However, all crystals grown from both conditions were long rod shaped and around 0.2-0.3mm in size. All crystals looked to be of a good enough quality to be used to get initial diffraction data.

Before any diffraction attempts were carried out, all crystals were soaked in mother liquor containing 10-20% (v/v) glycerol as a cryoprotectant and then mounted on a heated Goniometer head in a stream of liquid nitrogen using a Nikon Microscope and wire loop. Initial X-ray diffraction data was collected using a Rigaku RU-H3R rotating anode X-ray generator and a 300mm Mar image plate running on a Silicon Graphics workstation. Unfortunately, the diffraction data collected from this run only went up to a 7Å resolution. Since diffraction at this resolution would only be able to be used to determine protein domains and overall shape, it was decided that an alternative X-ray source was needed. However, it was possible from this data to calculate the space group to be P222 with cell dimensions a=89.1447Å, b=102.8808Å and c=198.1554Å.

A.



B.



C.

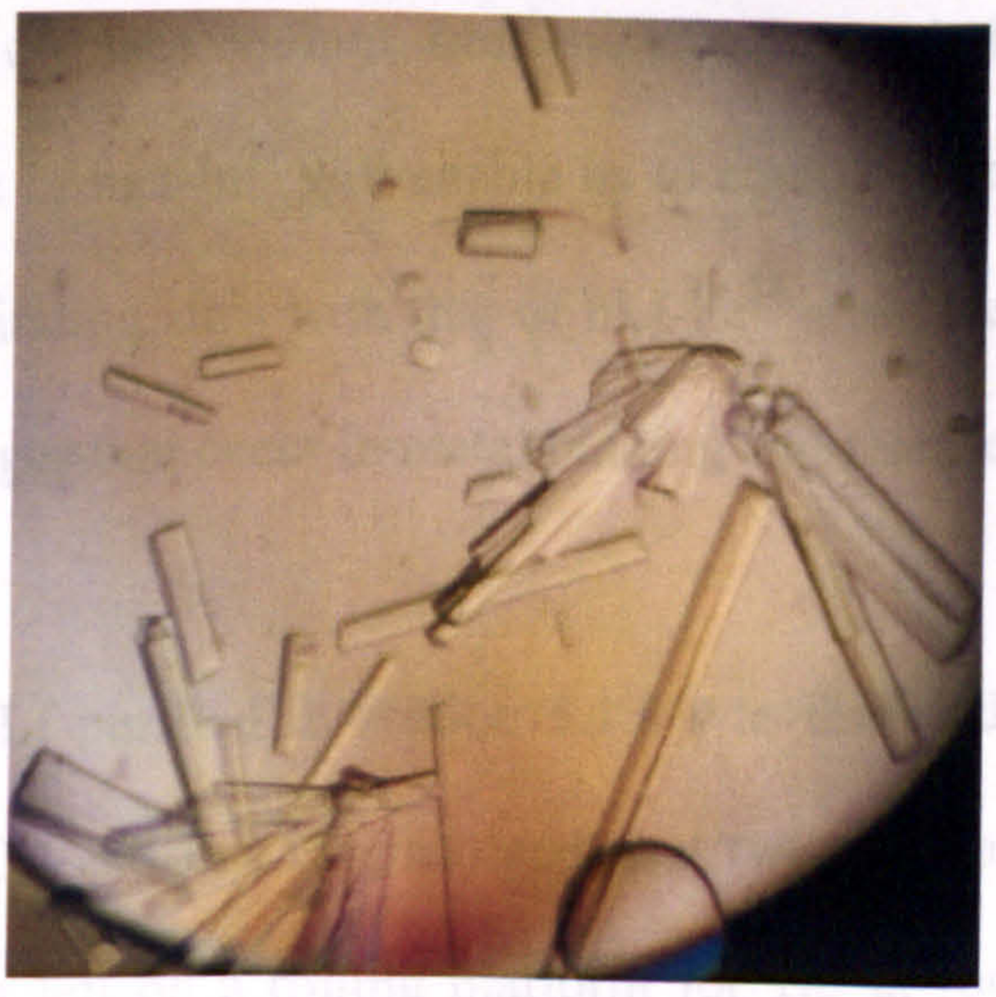


Figure 5.23. Examples of diferric rhSTf crystals grown in these studies. Crystals were visualised using a Leica MZFLIII microscope and photos taken using a mounted Carl Zeiss Axiocam.

In order to obtain higher resolutions, crystals were taken to the Synchrotron facility in Grenoble, France. Using the facilities available there, a 4Å resolution data set was collected, taking diffraction images every 1° for 90° (Figure 5.24). The data collected from all of these images was analysed using the CCP4 (Collaborative Computational Project Number 4) UNIX-based software suite. This software showed that there is likely to be two molecules of recombinant hSTf in each asymmetric unit and confirms the space group to be P222. Even though 4Å resolution is not good enough to solve the structure of hSTf by itself, molecular replacement analysis with rabbit serum transferrin (1JNF) and the N-lobe of hSTf (1N84) is currently under way.

5.2.3. Crystallisation of Native hLf-In^{3+}

This section discusses the screening of conditions for the crystallisation of hLf bound to indium(III), a metal shown by Battistuzzi *et al*, 1995 to bind to transferrin with an affinity on par with that of Iron(III). The reason for choosing indium(III) is due to its large atomic radii of 92pm, compared to iron(III)'s 67pm, the fact that iron does not have the ability to displace indium from transferrin, and due to its availability. Solving the structure of hLf-In^{3+} will enable us to compare the structural differences around both metal-binding sites with that of hLf-Fe^{3+} and give us an insight into the conformational changes that occur in metal binding and release.

Recombinant WT hLf purified in section 5.1.8 was dialysed into 20mM $NaHCO_3$ (pH 8 with 10M HCl) and In^{3+} (Sigma) added at twice the protein's molarity. This protein solution was left on a rolling platform for 1hr at room temperature before being desalted and freeze-dried using the protocol described in section 2.3.6.

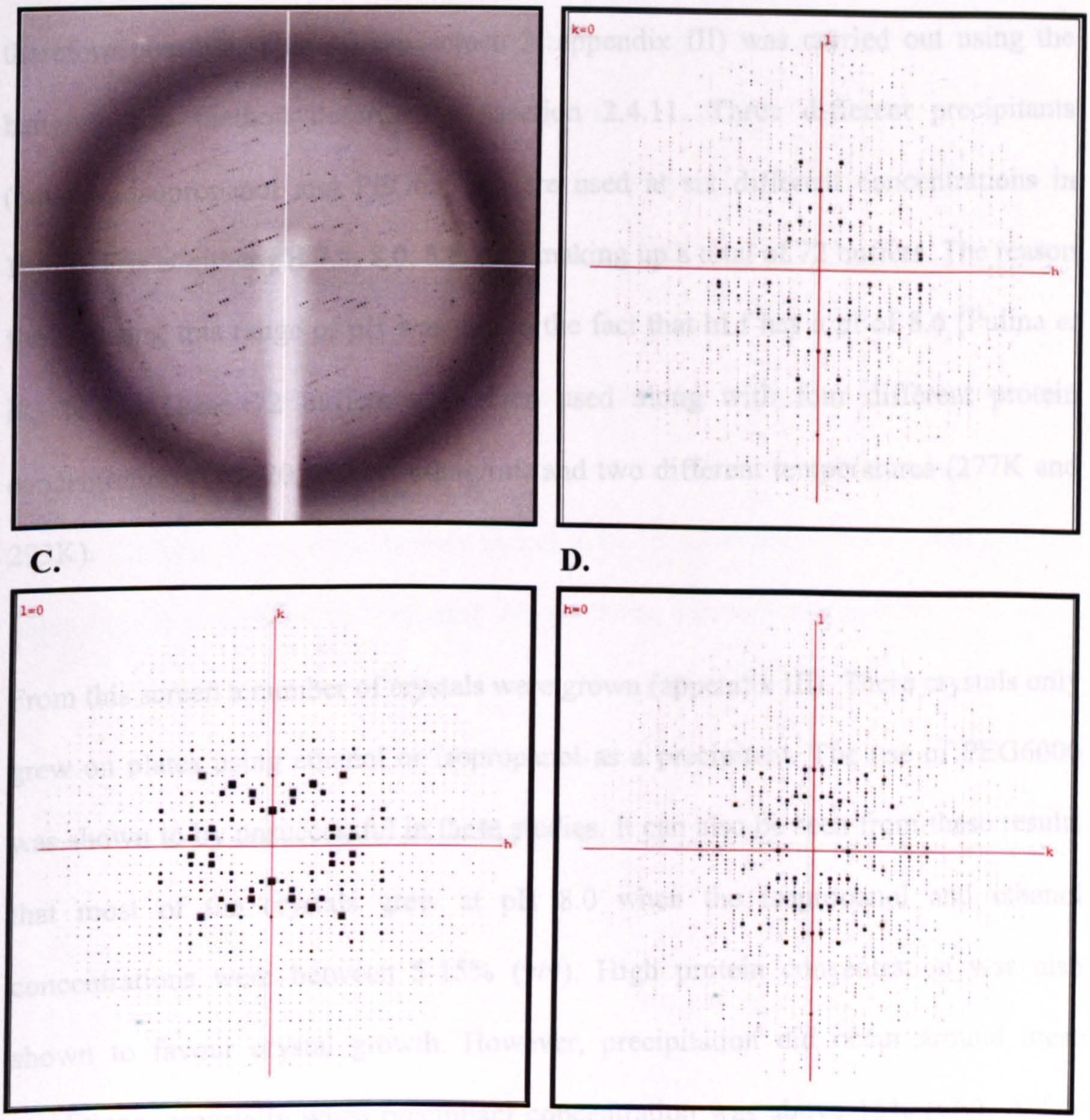


Figure 5.24. Examples of X-ray diffraction images of recombinant hSTf crystals using Synchrotron radiation (Grenoble, France). A: An example of an unprocessed diffraction image. B-D: diffraction images looking down the three axes (l, k and h) of the crystal.

As discussed before, hLf has been crystallised in its apo- and diferric form. A number of lactoferrins from other species have also been crystallised. Such species include, buffalo (Karthikeyan *et al*, 2000), bovine (Moore *et al*, 1997), equine (Kumar *et al*, 2002) and camel (Khan *et al*, 2001). From this published data, a relatively focused initial screen of conditions for the crystallisation of hLf-In^{3+} was therefore possible. This screen (screen 3, appendix III) was carried out using the hanging-drop method described in section 2.4.11. Three different precipitants (ethanol, isopropanol and PEG6000) were used at six different concentrations in 10mM Tris at either pH 7.5, 8.0, 8.5, or 9 making up a total of 72 buffers. The reason for choosing this range of pH was due to the fact that hLf has a pI of 8.6 (Pulina *et al*, 2002). These 72 buffers were then used along with four different protein concentrations (10, 20, 30 and 40mg/ml) and two different temperatures (277K and 293K).

From this screen a number of crystals were grown (appendix III). These crystals only grew on plates using ethanol or isopropanol as a precipitant. The use of PEG6000 was shown to by unsuccessful in these studies. It can also be seen from these results that most of the crystals grew at pH 8.0 when the isopropanol and ethanol concentrations were between 5-15% (v/v). High protein concentration was also shown to favour crystal growth. However, precipitation did occur around these conditions, especially when precipitant concentration was above 15% (v/v). A few crystals did grow outside these conditions, but these were usually small and irregular in shape.

Photographs of eight crystals grown in this screen are shown in Figure 5.25. All these crystals were of similar shape and ranged in size from 0.2-0.5mm in length. It is possible to see from this figure that there is a morphological difference between the crystals grown using ethanol as a precipitant rather than isopropanol. This difference is seen in the presence of striations in the isopropanol grown crystals. These striations are characteristic of crystal cracking. One possible reason for this could be that the drops became too dehydrated, causing the crystals to become unstable. It was concluded from this screen that the best hLf-In^{3+} crystals grew using buffer 8, 40mg/ml protein and at a temperature of 277K.

The collection of crystal diffraction data was attempted using an in-house rotating anode X-ray diffractor (45kV, 45mA, detector distance 200mm, 10min exposure per image). Unfortunately, crystals were found to be unstable and cracked easily during mounting. In an attempt to overcome this problem, a cryoprotectant was tried. This involved adding 1μl of 2-methyl-2,4-pentanediol (MPD) (Sigma) to the relative drop. The crystal was then passed through this solution before mounting. Again, crystals were found to be unstable and no diffraction data could be collected.

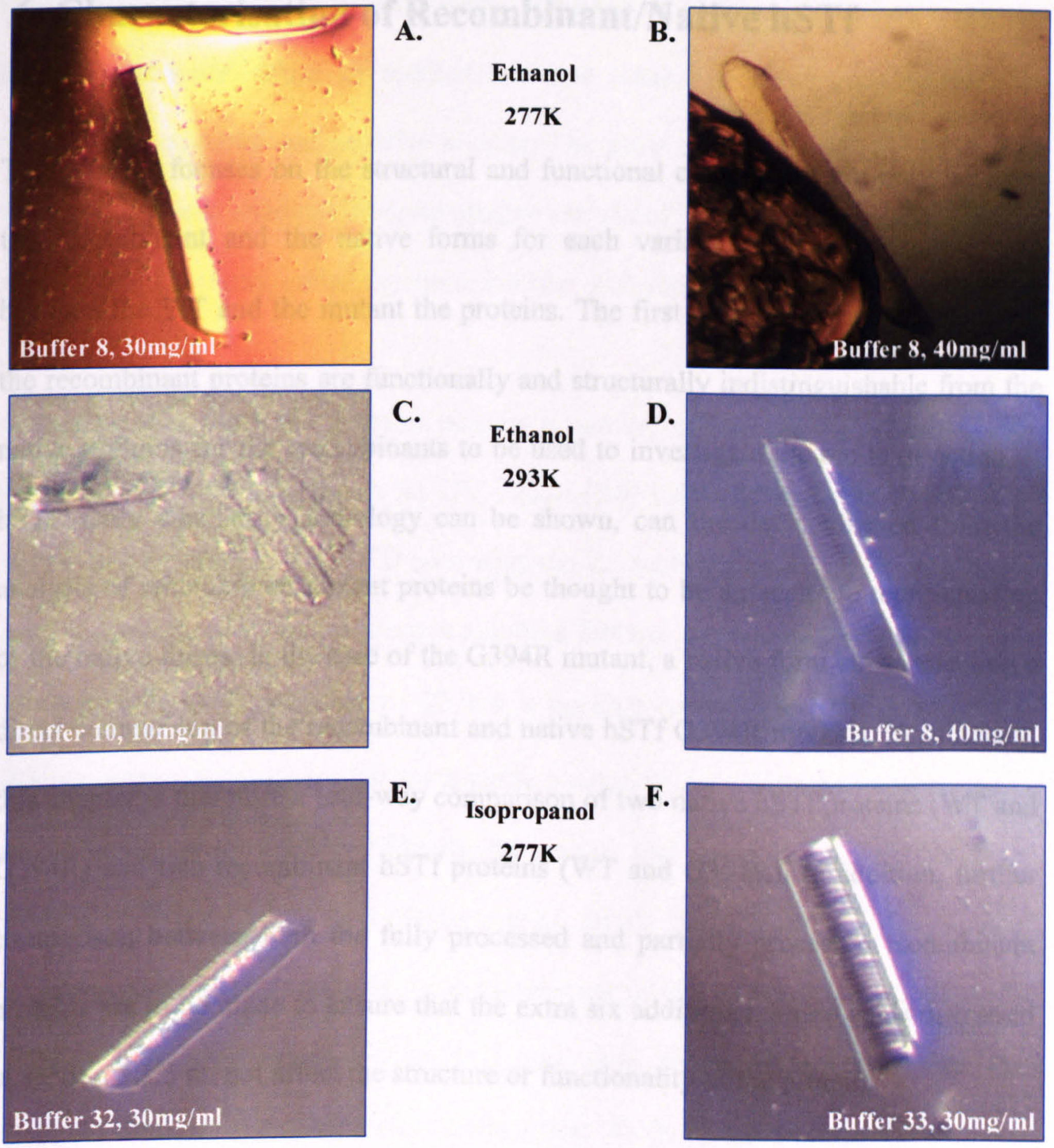


Figure 5.25. Examples of hLf-In^{3+} crystals grown in these studies. Crystals were visualised using a Leica MZFLIII microscope and photos taken using a mounted Carl Zeiss Axiocam.

6. Characterisation of Recombinant/Native hSTf

This chapter focuses on the structural and functional comparison not only between the recombinant and the native forms for each variant of the protein, but also between the WT and the mutant the proteins. The first comparison is to ensure that the recombinant proteins are functionally and structurally indistinguishable from the native proteins for the recombinants to be used to investigate the mode of action of hSTf. Only when true homology can be shown, can the data collected from the analysis of recombinant mutant proteins be thought to be an accurate representation of the native forms. In the case of the G394R mutant, a native form exists enabling a direct comparison of the recombinant and native hSTf G394R mutants. In summary, this chapter is therefore a four-way comparison of two native hSTf proteins (WT and G394R) and two recombinant hSTf proteins (WT and G394R). In addition, further comparison between both the fully processed and partially processed recombinant proteins are to be made to ensure that the extra six additional amino acids discussed in section 5.1.5 do not affect the structure or functionality of the protein.

6.1. Structural Characterisation

Four different techniques were used in this section to compare the structural characteristics of these four proteins: circular dichroism spectroscopy (CD), electron paramagnetic resonance (EPR), dynamic light scattering and immunodiffusion.

6.1.1. Circular Dichroism Spectroscopy

Circular dichroism spectroscopy measures differences in the absorption of left-handed circularly polarized light versus right-handed circularly polarized light,

which arise due to structural asymmetry. The absence of regular structure results in zero CD intensity, while an ordered structure results in a spectrum, which can contain both positive and negative signals. The most commonly used units of CD intensity are mean residue ellipticity (degree cm^2 $dmol^{-1}$) and the difference in molar extinction coefficients, called the molar circular dichroism ($\Delta\varepsilon$) (M^{-1} cm^{-1}).

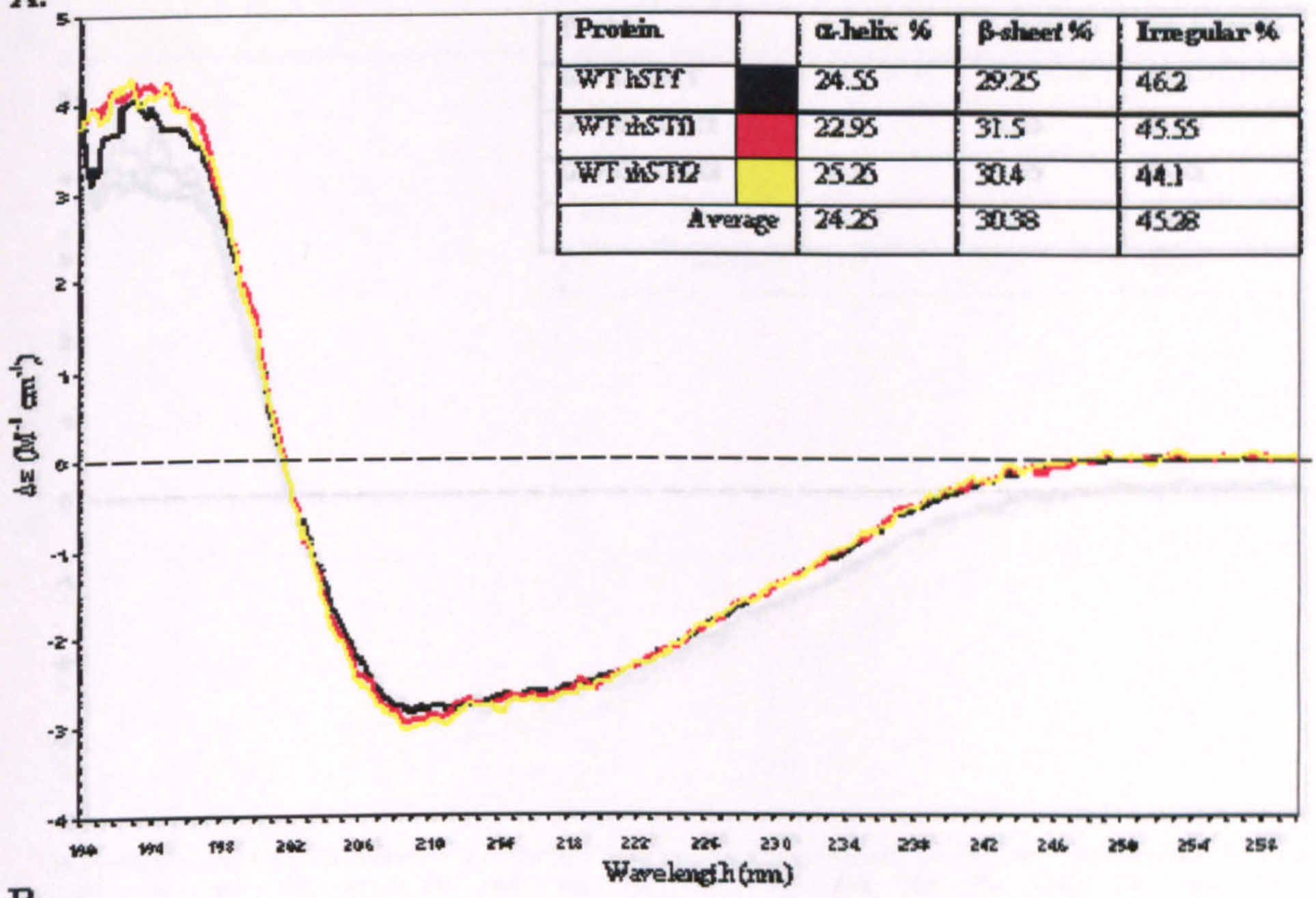
In relation to proteomics, proteins possess a number of chromophores, which can give rise to CD signals. In the far UV region (240-180nm), which corresponds to peptide bond absorption, the CD spectrum can be analysed using reference data to give the content of regular secondary structure features such as α-helix and β-sheet. For an α-helix, a negative $\Delta\varepsilon$ at ~222nm, a positive $\Delta\varepsilon$ ~190 and a negative $\Delta\varepsilon$ at ~208nm occurs. β-sheets are known to cause a negative $\Delta\varepsilon$ ~ 215nm, a positive $\Delta\varepsilon$ ~ 198 and a negative $\Delta\varepsilon$ ~175nm. Finally, for random coil, a negative $\Delta\varepsilon$ ~212 nm and a negative $\Delta\varepsilon$ ~ 195 nm is seen (Wallace *et al*, 2001). The CD spectrum in the near UV region (320-260nm) reflects the environments of the aromatic amino acid side chains and the presence of disulphide bonds, thus giving information about the tertiary structure of the protein (Kelly and Price, 2000). More specifically, the near UV absorption of the disulphide bond occurs near 260nm, tryptophan (290 and 305nm), tyrosine (275 and 283nm) and phenylalanine (255 and 270nm). However, like all spectroscopic techniques, the CD signal reflects an average of the entire molecular population. It cannot determine which specific residues are involved in certain structures. CD is therefore a very powerful tool in proteomics and is particularly suited for such studies. In this section, CD is used as a 'fingerprint' to compare all the different forms of protein.

Using the protocol described in the section 2.4.1, apo and diferric preparations of the six different transferrins, including the recombinant isoforms, purified in the previous chapter were analysed in both the far and near UV range. Apo and diferric preparations were made following the protocol described in section 2.3.5.

Figure 6.1, 6.2, 6.3 and 6.4 show the spectra for both CD and UV from 190 - 260nm, and the relative α-helix, β-sheet and irregular secondary structure percentages calculated from data averaged from two scans per protein. Results obtained indicate that the recombinant forms of both variant proteins produce spectra which are superimposable to the ones obtained with their respective native forms, and for both apo and diferric forms. This indicates, along with the percentage of secondary structure calculated, that the four recombinant proteins are correctly folded and show minor changes in absorption in the near UV.

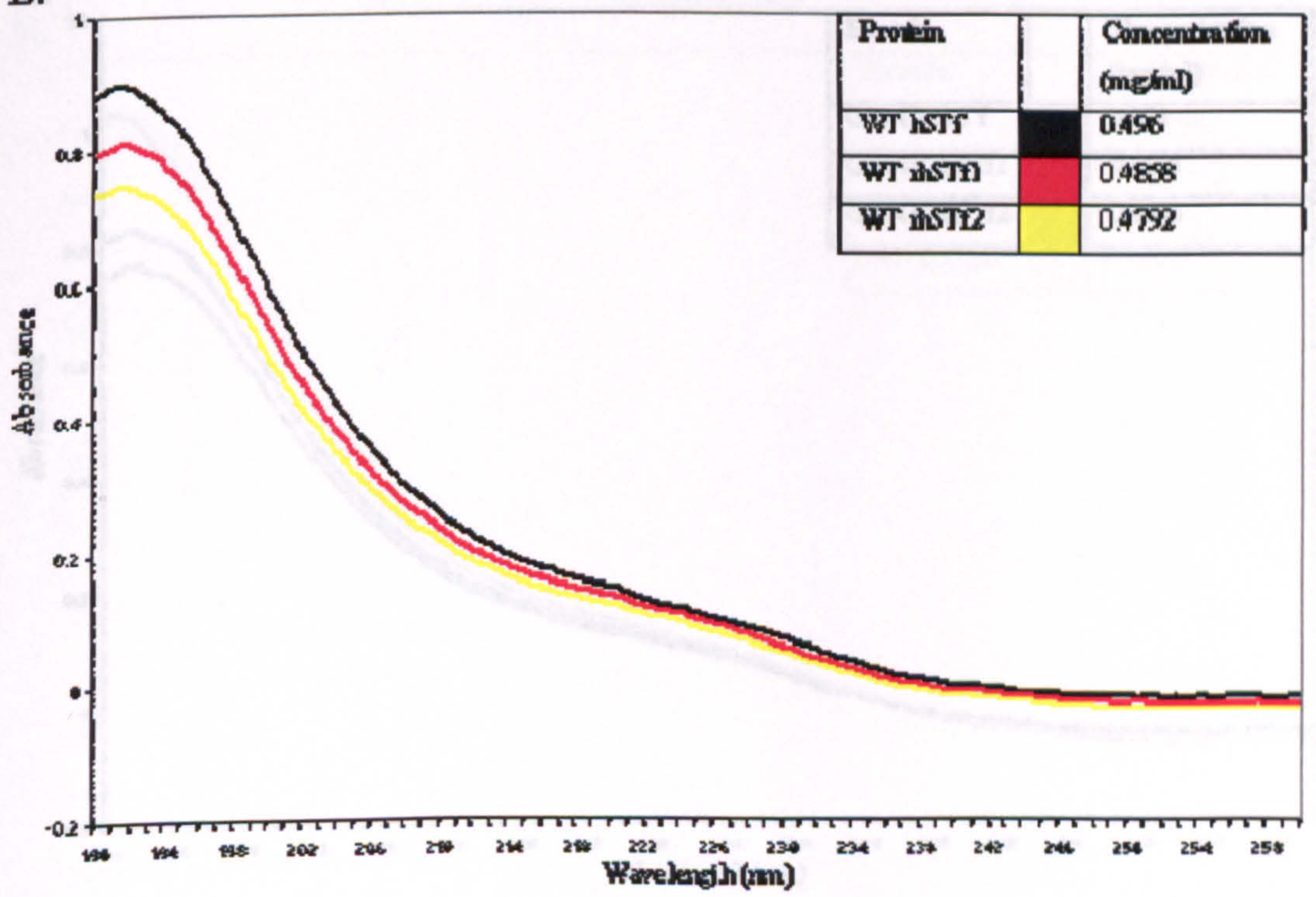
In order to look at the overall differences in secondary structure between the apo and diferric forms of the WT and G394R proteins, each data set from both variants were averaged (Figure 6.5). These results show that both the WT and G394R hSTf produce very similar CD spectra in the far UV range regardless of their iron content. This similarity can also be seen in the variation of percentage secondary structure calculated (v). These secondary structure percentages only vary by $^{+}/_{-}$ ~1% from the average. In terms of the number of residues (equation 1), this variation is approximately $^{+}/_{-}$ ~2 residues, out of a total of 679 residues. From this, it can therefore be seen that this variation is probably not significant and more likely to be due to experimental error rather than resulting from structural differences between

A.



Protein		α-helix %	β-sheet %	Irregular %
WT hSTf		24.55	29.25	46.2
WT rhSTf1		22.95	31.5	45.55
WT rhSTf2		25.25	30.4	44.1
Average		24.25	30.38	45.28

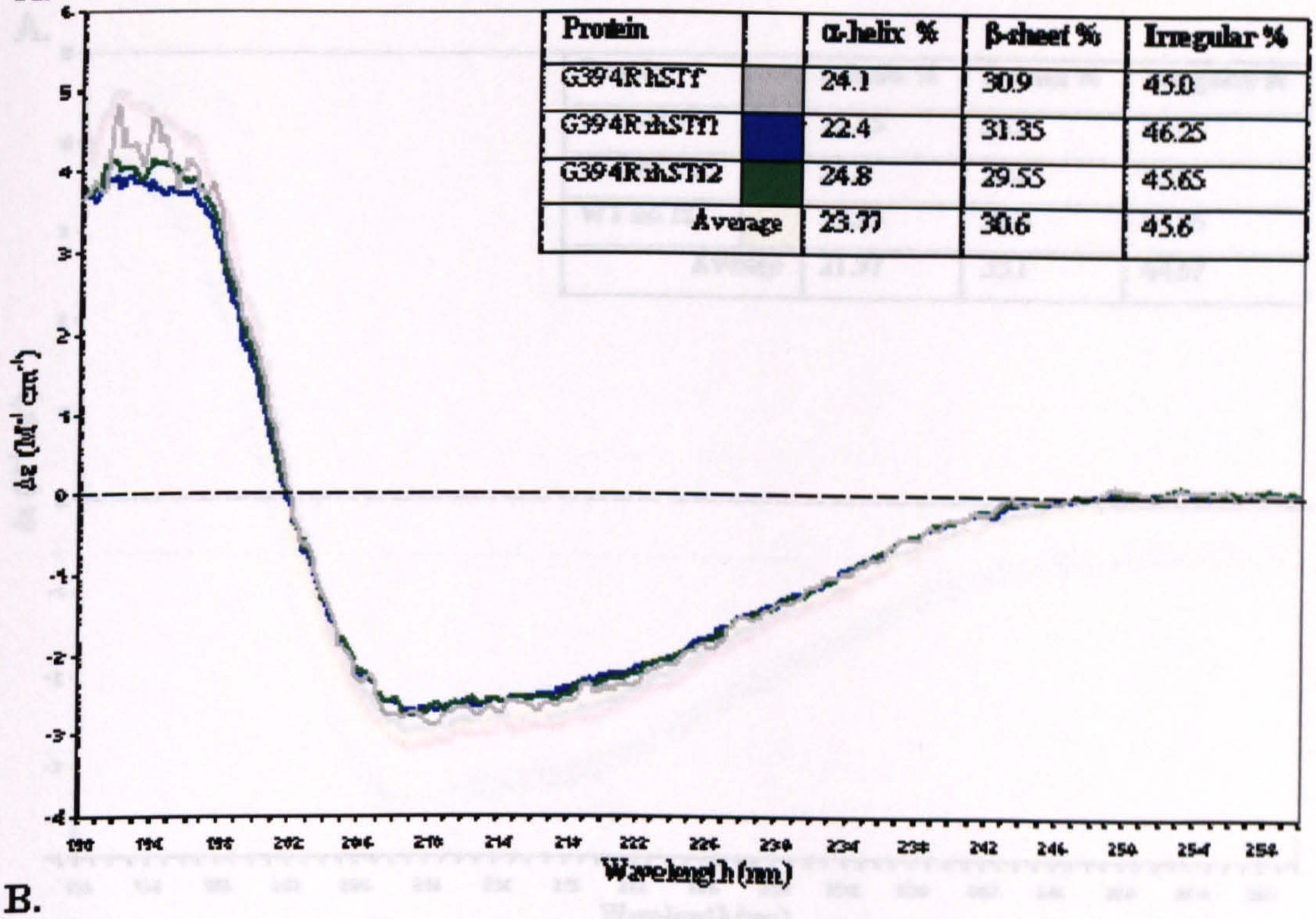
B.



Protein		Concentration (mg/ml)
WT hSTf		0.496
WT rhSTf1		0.4858
WT rhSTf2		0.4792

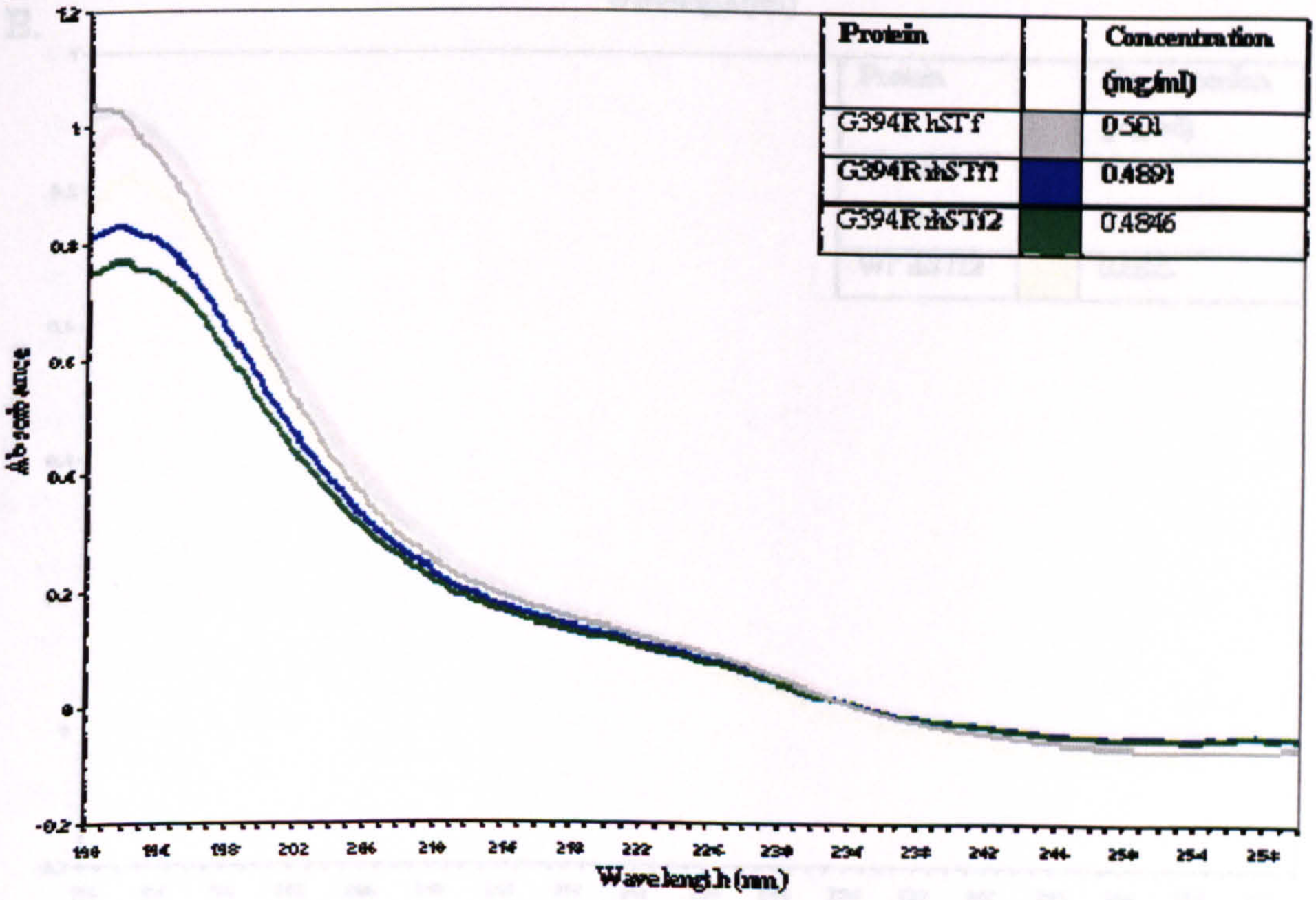
Figure 6.1. Far-UV (190-260nm) CD (A) and UV (B) spectra of native and recombinant WT apo-hSTf. All scans were carried out in duplicate with the average shown above. Black line: Native WT hSTf, red line: partially processed WT recombinant hSTf and yellow line: fully processed WT recombinant hSTf.

A.



Protein		α-helix %	β-sheet %	Irregular %
G394R hSTf		24.1	30.9	45.0
G394R rhSTf1		22.4	31.35	46.25
G394R rhSTf2		24.8	29.55	45.65
Average		23.77	30.6	45.6

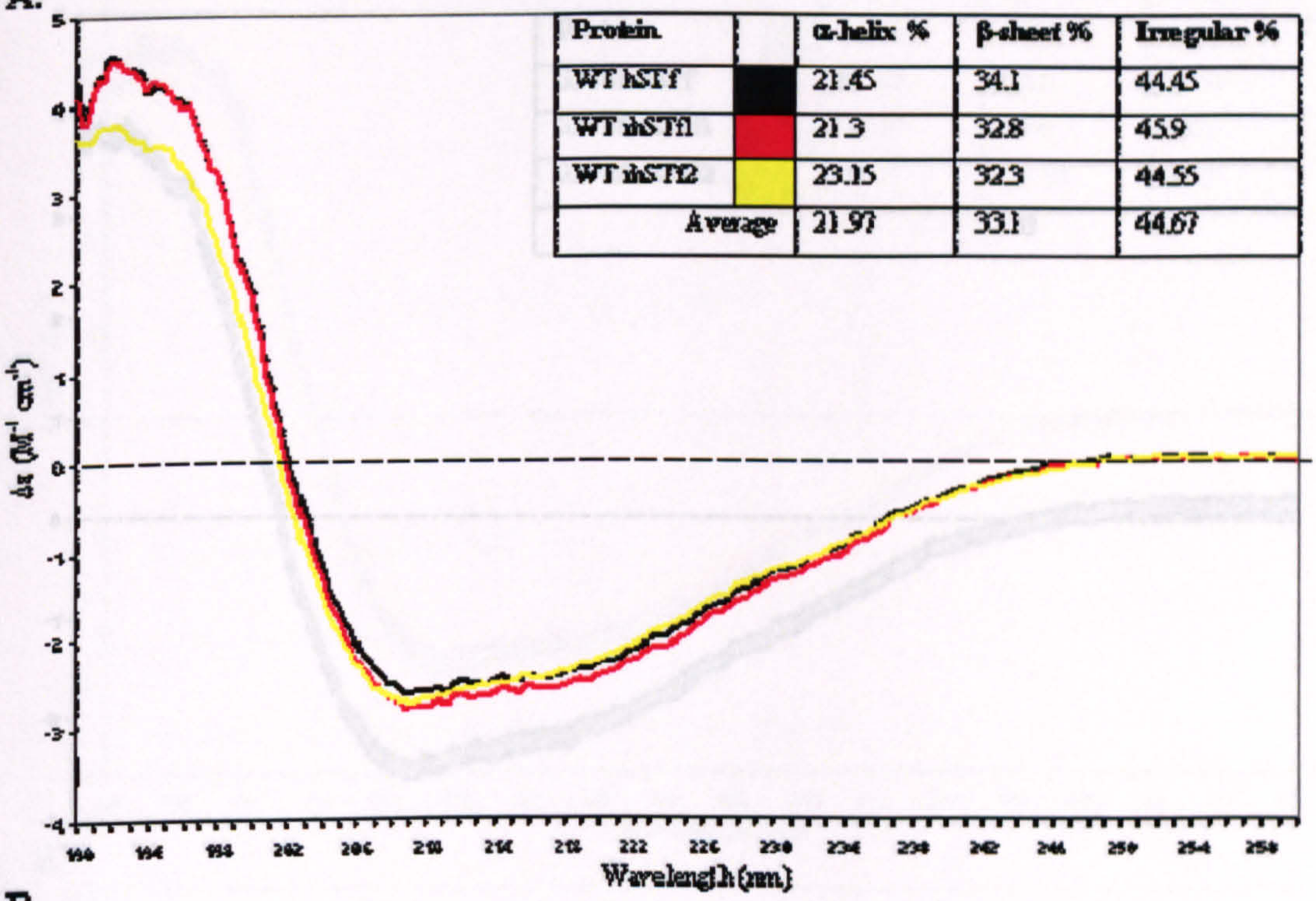
B.



Protein		Concentration (mg/ml)
G394R hSTf		0.501
G394R rhSTf1		0.4891
G394R rhSTf2		0.4846

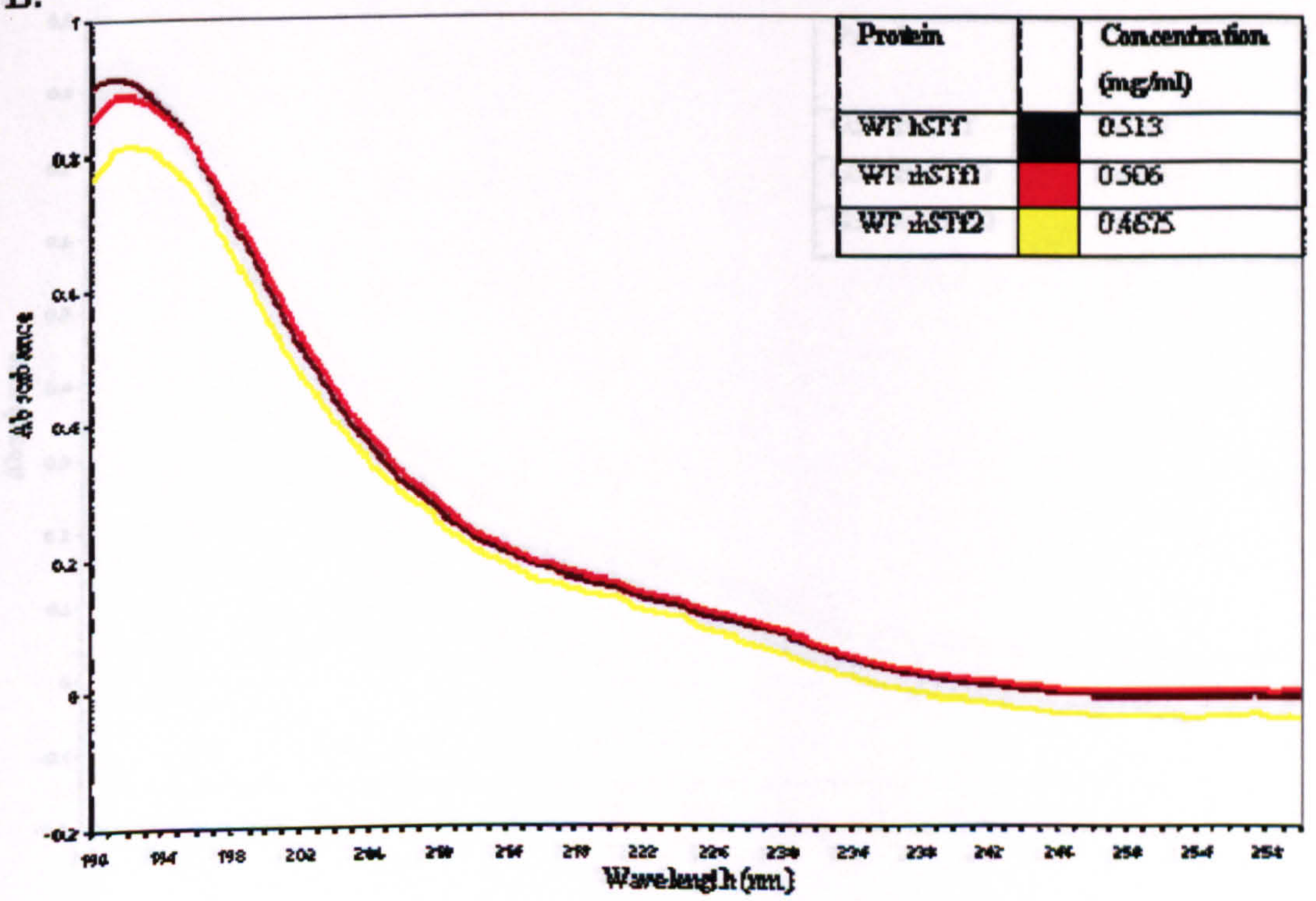
Figure 6.2. Far-UV (190-260nm) CD (A) and UV (B) spectra of native and recombinant G394R apo-hSTf. All scans were carried out in duplicate with the average shown above. Grey line: Native G394R hSTf, blue line: partially processed G394R recombinant hSTf and green line: fully processed G394R recombinant protein.

A.



Protein		α-helix %	β-sheet %	Irregular %
WT hSTf		21.45	34.1	44.45
WT rhSTf1		21.3	32.8	45.9
WT rhSTf2		23.15	32.3	44.55
Average		21.97	33.1	44.67

B.



Protein		Concentration (mg/ml)
WT hSTf		0.513
WT rhSTf1		0.506
WT rhSTf2		0.4675

Figure 6.3. Far-UV (190-260nm) CD (A) and UV (B) spectra of native and recombinant WT diferric-hSTf. All scans were carried out in duplicate with the average shown above. Black line: Native WT hSTf, red line: partially processed WT recombinant hSTf and yellow line: fully processed WT recombinant hSTf.

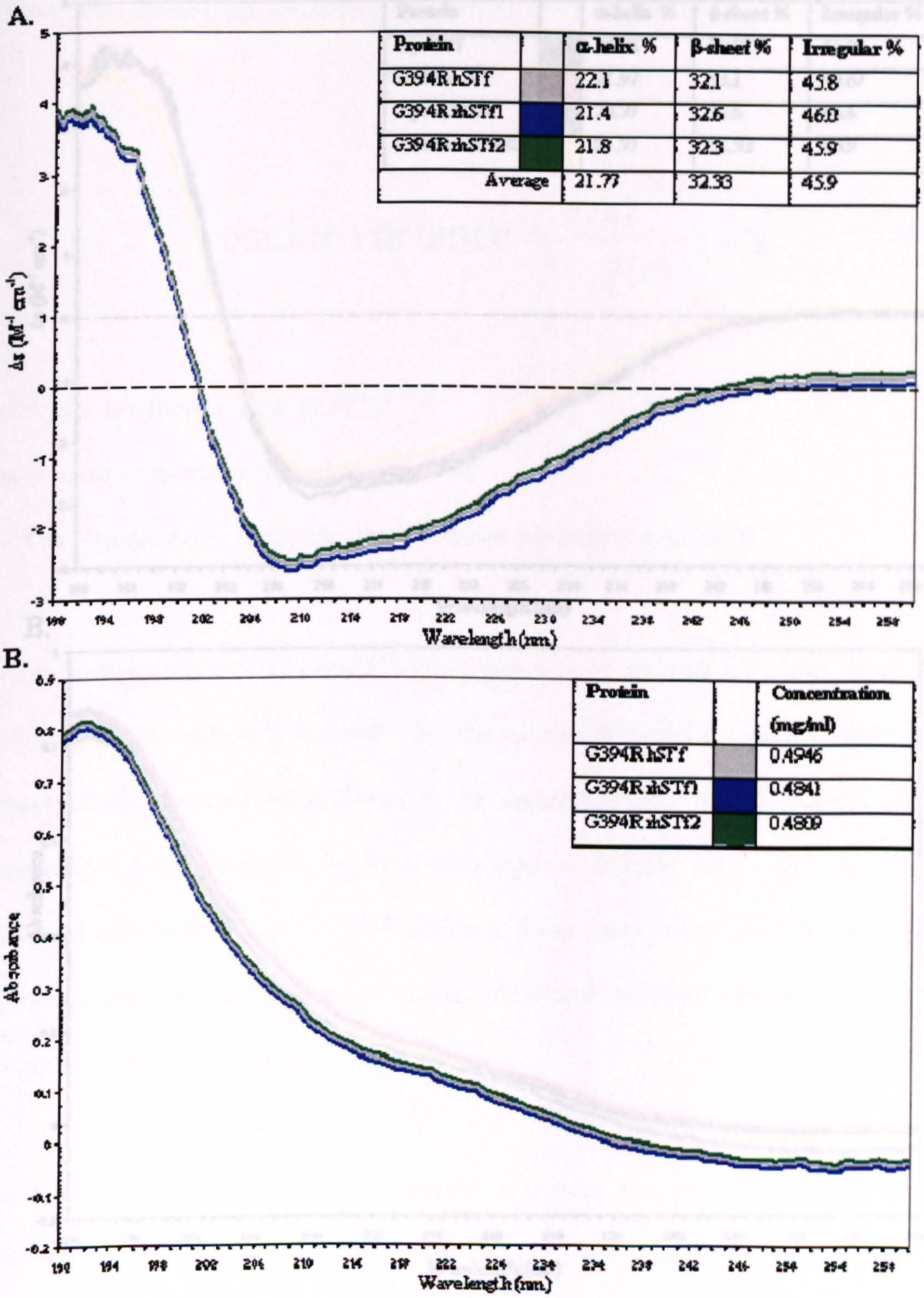
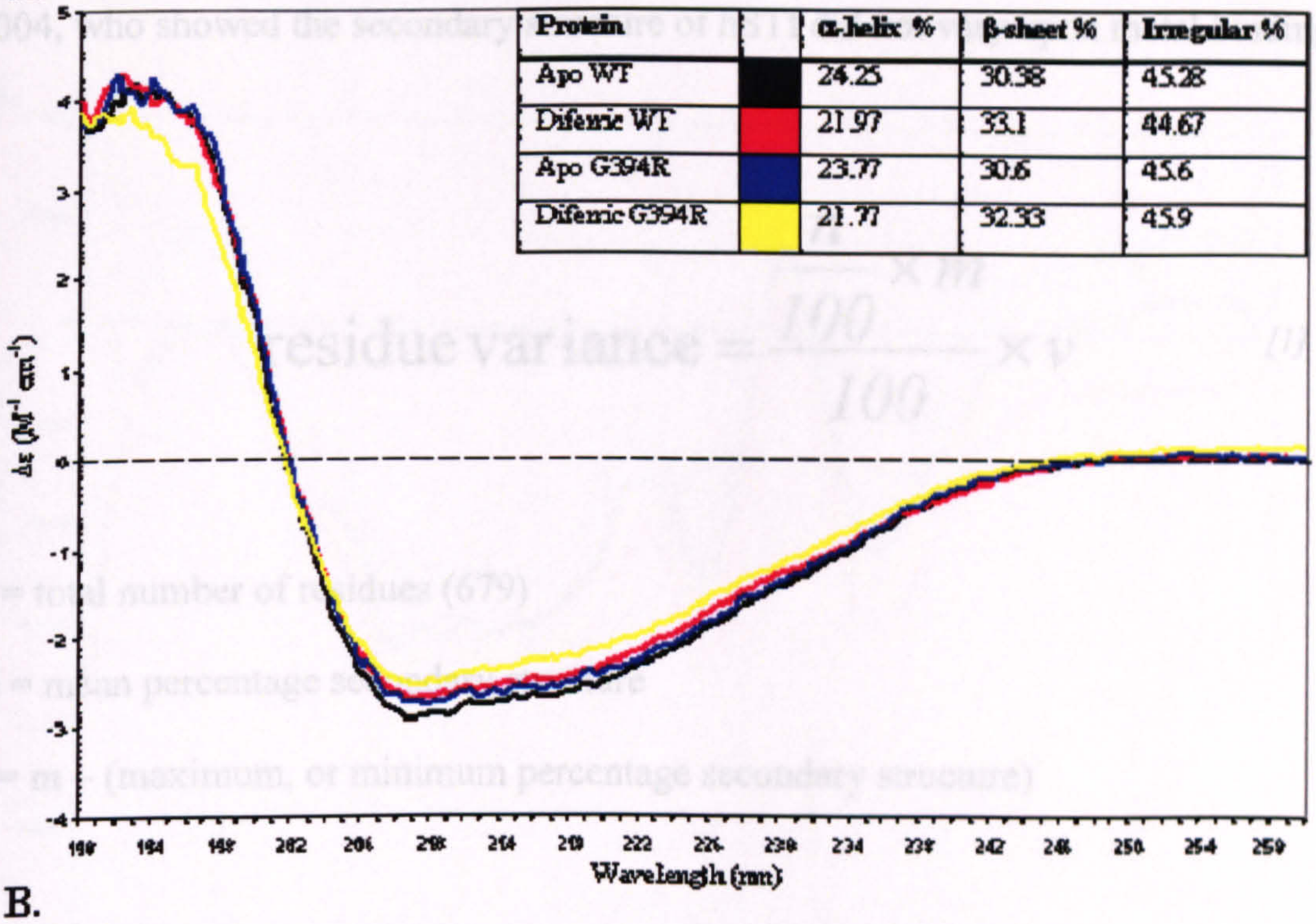


Figure 6.4. Far-UV (190-260nm) CD (A) and UV (B) spectra of native and recombinant G394R diferric-hSTf. All scans were carried out in duplicate with the average shown above. Grey line: Native G394R hSTf, blue line: partially processed G394R recombinant hSTf and green line: fully processed G394R recombinant protein.

A.

Protein		α-helix %	β-sheet %	Irregular %
Apo WT		24.25	30.38	45.28
Diferric WT		21.97	33.1	44.67
Apo G394R		23.77	30.6	45.6
Diferric G394R		21.77	32.33	45.9



B.

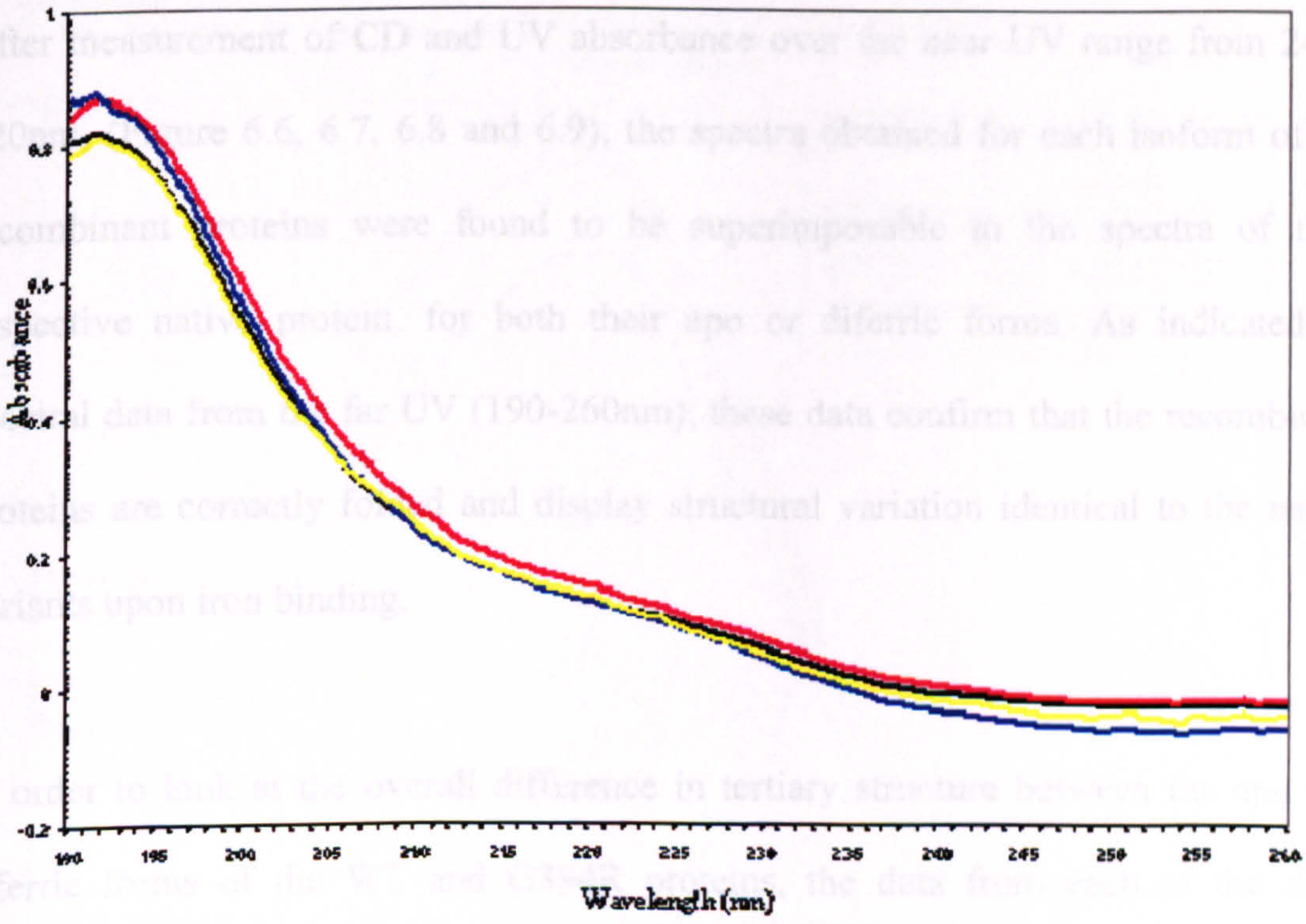


Figure 6.5. Far-UV (190-260nm) CD spectra (CD) and Near-UV (249-320) UV spectra (B) of WT and G394R hSTf, in both apo and diferric forms. Data compiled from averaged data from all six proteins. Black line: apo hSTf WT, Red line: diferric hSTf WT, blue line: apo hSTf G394R and Yellow line: diferric hSTf G394R.

the four proteins. This conclusion is consistent with the observation of Zhang *et al*, 2004, who showed the secondary structure of hSTf did not vary upon metal binding.

$$\text{residue variance} = \frac{\frac{n}{100} \times m}{100} \times v \qquad [1]$$

n = total number of residues (679)

m = mean percentage secondary structure

v = **m** – (maximum, or minimum percentage secondary structure)

After measurement of CD and UV absorbance over the near UV range from 249 - 320nm, (Figure 6.6, 6.7, 6.8 and 6.9), the spectra obtained for each isoform of the recombinant proteins were found to be superimposable to the spectra of their respective native protein, for both their apo or diferric forms. As indicated by spectral data from the far UV (190-260nm), these data confirm that the recombinant proteins are correctly folded and display structural variation identical to the native variants upon iron binding.

In order to look at the overall difference in tertiary structure between the apo and diferric forms of the WT and G394R proteins, the data from each of the three proteins, for both variants, were averaged (Figure 6.10). Although identical spectra were obtained for both apo proteins, considerable differences can be seen after iron binding, not only between the spectra obtained for the apo and the holo-forms, but also between both spectra obtained for both holo-forms of the variants.

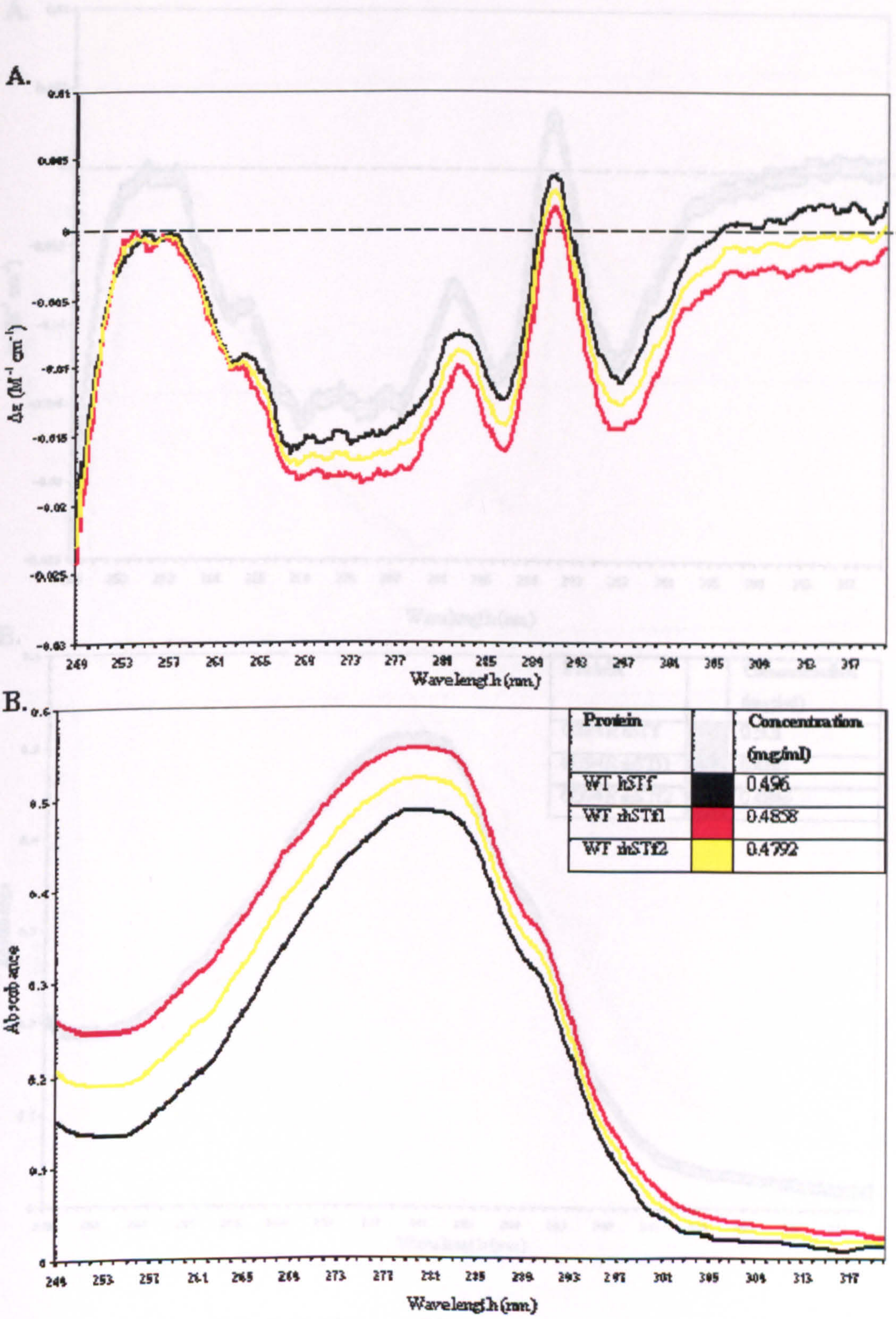


Figure 6.6. Near-UV (249-320nm) CD (A) and UV (B) spectra of native and recombinant WT apo-hSTf. All scans were carried out in duplicate with the average shown above. Black line: Native WT hSTf, red line: partially processed WT recombinant hSTf and yellow line: fully processed WT recombinant protein.

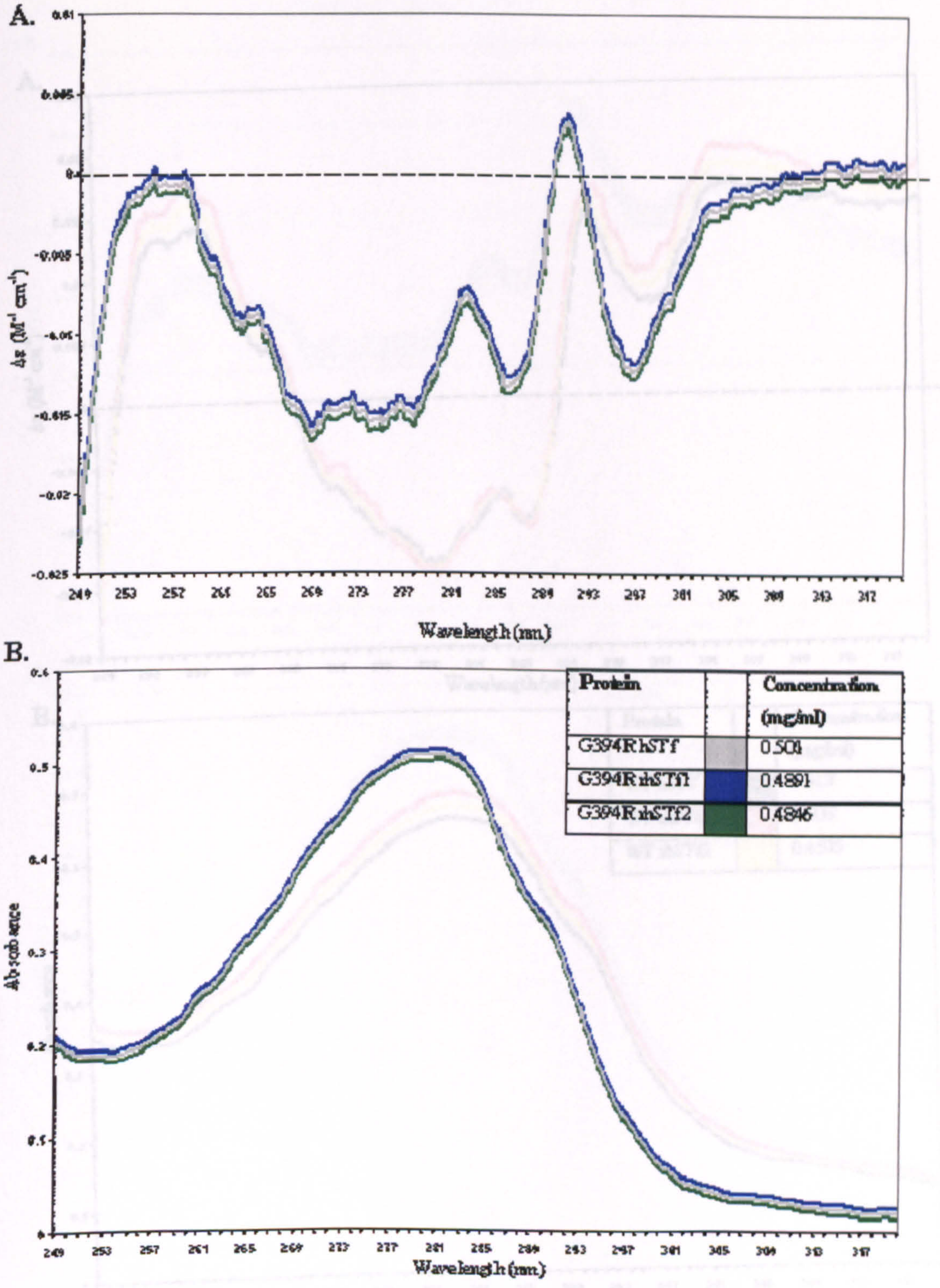


Figure 6.7. Near-UV (249-320nm) CD (A) and UV (B) spectra of native and recombinant G394R apo-hSTf. All scans were carried out in duplicate with the average shown above. Grey line: Native G394R hSTf, blue line: partially processed G394R recombinant hSTf and green line: fully processed G394R recombinant protein.

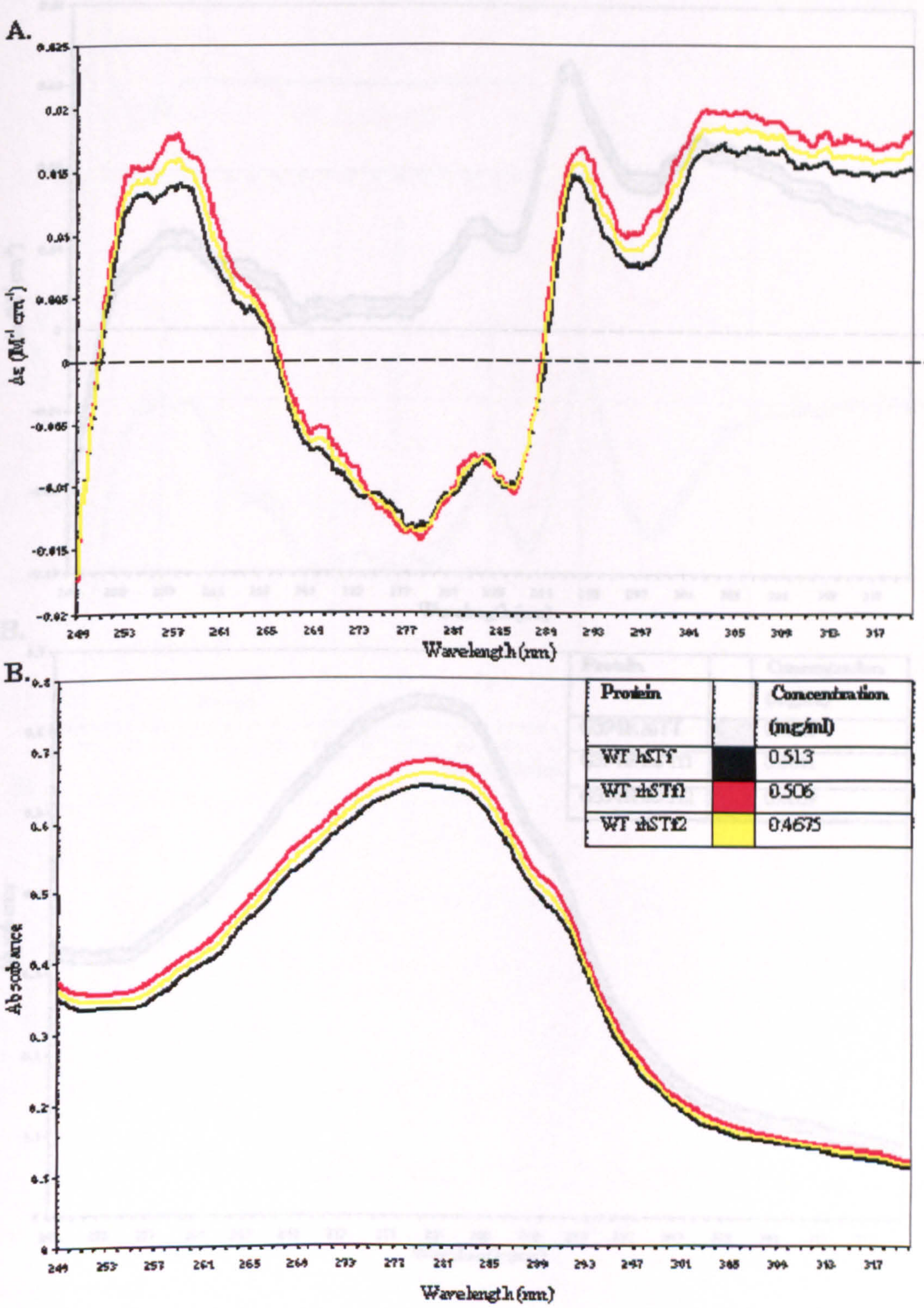


Figure 6.8. Near-UV (249-320nm) CD (A) and UV (B) spectra of native and recombinant WT diferric-hSTf. All scans were carried out in duplicate with the average shown above. Black line: Native WT hSTf, red line: partially processed WT recombinant hSTf and yellow line: fully processed WT recombinant protein.

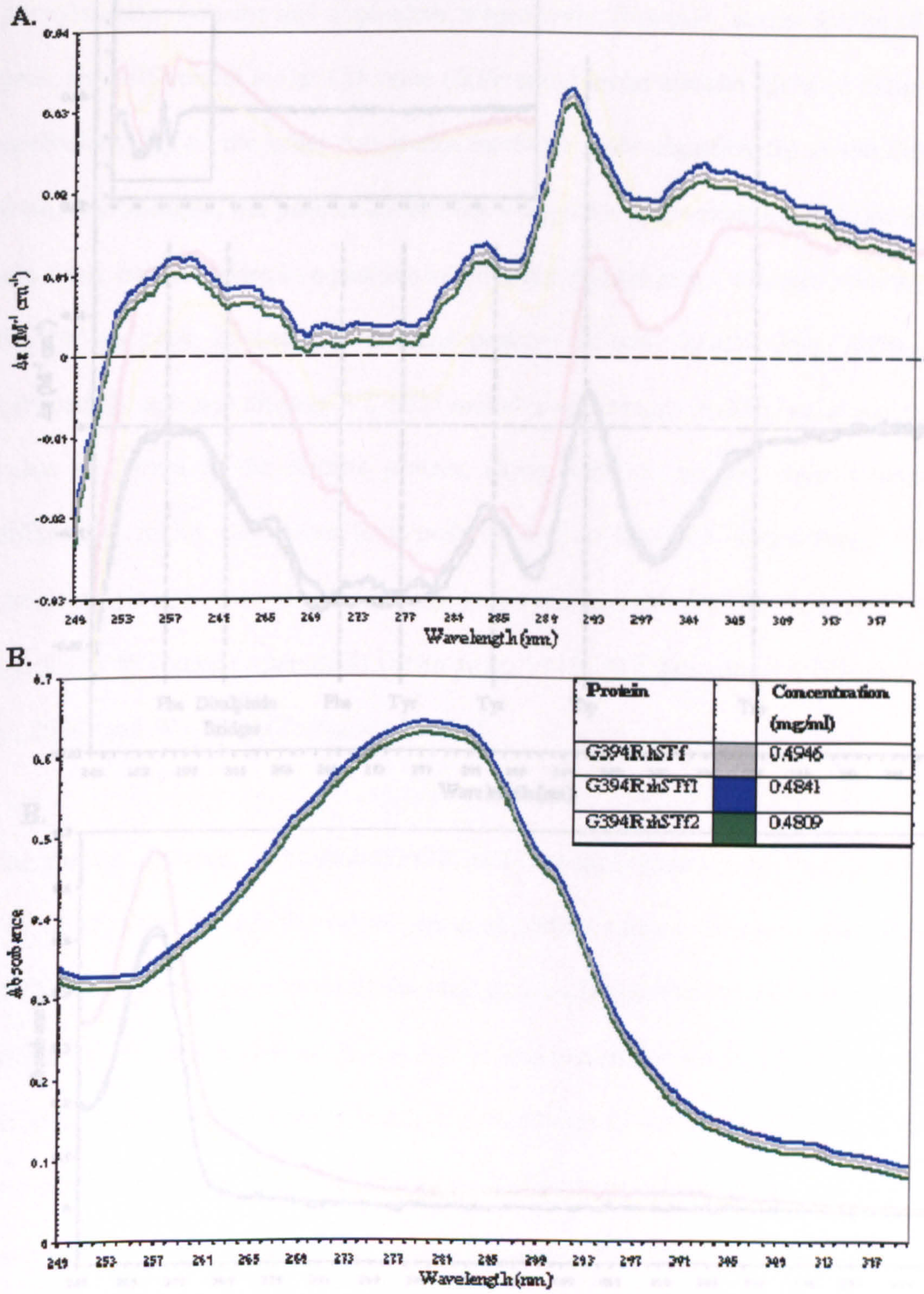


Figure 6.9. Near-UV (249-320nm) CD (A) and UV (B) spectra of native and recombinant G394R diferric-hSTf. All scans were carried out in duplicate with the average shown above. Grey line: Native G394R hSTf, blue line: partially processed G394R recombinant hSTf and green line: fully processed G394R recombinant protein.

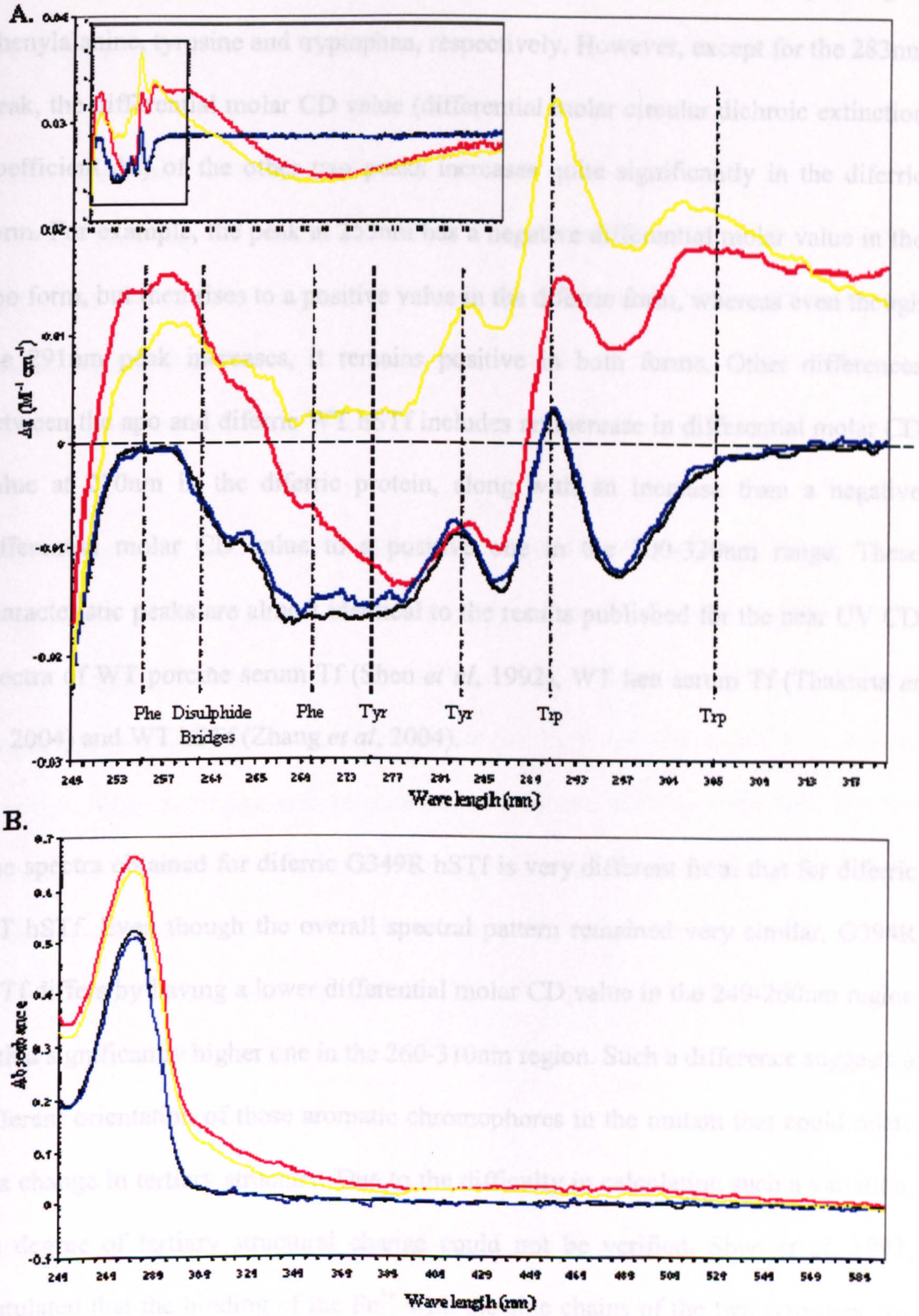


Figure 6.10. Near-UV (249-320nm) CD spectra (A) taken from the inset scan (249-600nm) and Near-UV (249-600nm) UV spectra (B) of WT and G394R hSTf, in both apo and diferric forms. Data compiled from averaged data from all six proteins. Black line: apo hSTf WT, Red line: diferric hSTf WT, blue line: apo hSTf G394R and Yellow line: diferric hSTf G394R.

For WT hSTf, both forms show peaks at 253, 283 and 291nm, corresponding to phenylalanine, tyrosine and tryptophan, respectively. However, except for the 283nm peak, the differential molar CD value (differential molar circular dichroic extinction coefficient $\Delta\varepsilon$) of the other two peaks increases quite significantly in the diferric form. For example, the peak at 253nm has a negative differential molar value in the apo form, but then rises to a positive value in the diferric form, whereas even though the 291nm peak increases, it remains positive in both forms. Other differences between the apo and diferric WT hSTf includes an increase in differential molar CD value at 270nm in the diferric protein, along with an increase from a negative differential molar CD value to a positive one in the 300-320nm range. These characteristic peaks are almost identical to the results published for the near UV CD spectra of WT porcine serum Tf (Shen *et al*, 1992), WT hen serum Tf (Thakurta *et al*, 2004) and WT hSTf (Zhang *et al*, 2004).

The spectra obtained for diferric G349R hSTf is very different from that for diferric WT hSTf. Even though the overall spectral pattern remained very similar, G394R hSTf differs by having a lower differential molar CD value in the 249-260nm region with a significantly higher one in the 260-310nm region. Such a difference suggests a different orientation of these aromatic chromophores in the mutant that could relate to a change in tertiary structure. Due to the difficulty in calculating such a variation, the degree of tertiary structural change could not be verified. Shen *et al*, 1992, postulated that the binding of the Fe^{3+} with the side chains of the two tyrosines, the histidine and the aspartate could significantly affect the absorbance of these tyrosines as well as the neighbouring tryptophan residues. Therefore, rather than a direct major structural alteration due to the presence of the mutation G394R, the difference

between both spectra could be the indirect result of fine structural differences around the iron-binding site.

6.1.2. Electron Paramagnetic Resonance

Electron paramagnetic resonance (EPR), also known as electron spin resonance (ESR), is the name given to the process of resonant absorption of microwave radiation by paramagnetic ions or molecules, with at least one unpaired electron spin, and in the presence of a magnetic field.

Electrons possess spin magnetic moments that can assume two quantum mechanical orientations. In electron pairs, the net moment is zero and thus the application of an external magnetic field will not affect electron spin. However, in compounds with unpaired electrons, such as transition metals (i.e. iron) and free-radicals, the electron has the ability to change spin orientation, or resonate, between spins with moments parallel or anti-parallel to an externally applied magnetic field. Transition between electron spin states, of which only two are allowed, according to Planck's law involves the absorption of a comparatively small amount of energy (when $\Delta E = h\nu$, where h is Planck's constant (6.626×10^{-34}) and ν is the frequency of the radiation).

In EPR, an unpaired electron is held in a specific spin orientation by the application of a magnetic field. Energy, in the form of microwave radiation, is then applied to the system. In relation to the Bruker EPR spectrometer used in these studies, this microwave radiation is kept at a constant frequency ($h\nu$) (9.38GHz) whilst the magnetic field strength (B_0) is then scanned until transition between electron spin states occurs. The strength of magnetic field required to do this is known as the field

for resonance. However, this field for resonance is not a unique fingerprint for identification of a compound because spectra can be acquired at several different frequencies. In order to overcome this, the g-factor is used. This g-factor is a proportionality constant approximately equal to 2 for most samples, but which varies depending on the electronic configuration of the radical or ion. The g-factor was calculated using the following equation (2), where μ_B is the Bohr Magneton ($9.2740154 \times 10^{-24}$).

$$g = \frac{h\nu}{\mu_B B_0} \qquad [2]$$

Measurements of the g-factor provide some useful information, but do not reveal very much about the molecular structure of our sample. However, when paramagnetic ions are present within molecules, a number of factors can influence the transition between electron spin states once a magnetic field has been applied. The magnetic moment of these electrons can be influenced by the molecules own magnetic field, or by the magnetic field of surrounding molecules, or even by interactions with nearby spinning nuclei (known as hyperfine interactions). These influences manifest themselves in a change in the g-value and are useful for the identification of the source of resonance.

For molecules such as transferrin, the task of quantitative spectra interpretation is an immense and highly problematic one. Therefore, the purpose of EPR in these studies was more of a qualitative tool, to compare the environments surrounding the Fe^{3+} ions in each of the transferrins under investigation.

Figure 6.11 shows a comparison of EPR spectra for both WT diferric proteins and Figure 6.12 a comparison of EPR spectra for both the G394R diferric proteins. All proteins were made diferric following the protocol described in section 2.3.5. The two recombinants used were from the fully processed fractions (section 5.1.5). The superimposable spectra obtained for both WT proteins are indicative of an identical chemical environment around the ferric iron for both native and recombinant proteins, with a resonance occurring at a g-value of 4.3, which is consistent with the g-value shown for the typical high-spin ferric iron-bicarbonate complex in WT hSTf (Pinkowitz and Aisen, 1972). Similarly, the EPR spectra obtained for the native and recombinant G394R proteins, are also both almost identical, with the resonance still fixed at a g-value around 4.3, and indicate therefore an identical chemical environment around the ferric iron.

When comparing the WT protein spectrum to the G394R mutant protein spectrum, a number of differences can be observed (Figure 6.13). These differences include a reduction in size of peak *a* in the G394R protein, along with the lowering of a shelf in the spectrum at point *b*, 155mT. When comparing the spectrum of this G394R mutant to the published hSTf spectrum, a close resemblance to the published spectra produced for the N-monoferric transferrin (Aisen *et al,* 1978) can be observed. This suggests that the positively charged side-chain of R394 is either interacting directly in some way with the ferric ion in the C-lobe, or indirectly perturbing the iron binding site. Either way, this indicates that the normal rhombic symmetry of the bound iron has been affected. Other differences can also be noticed in the magnetic

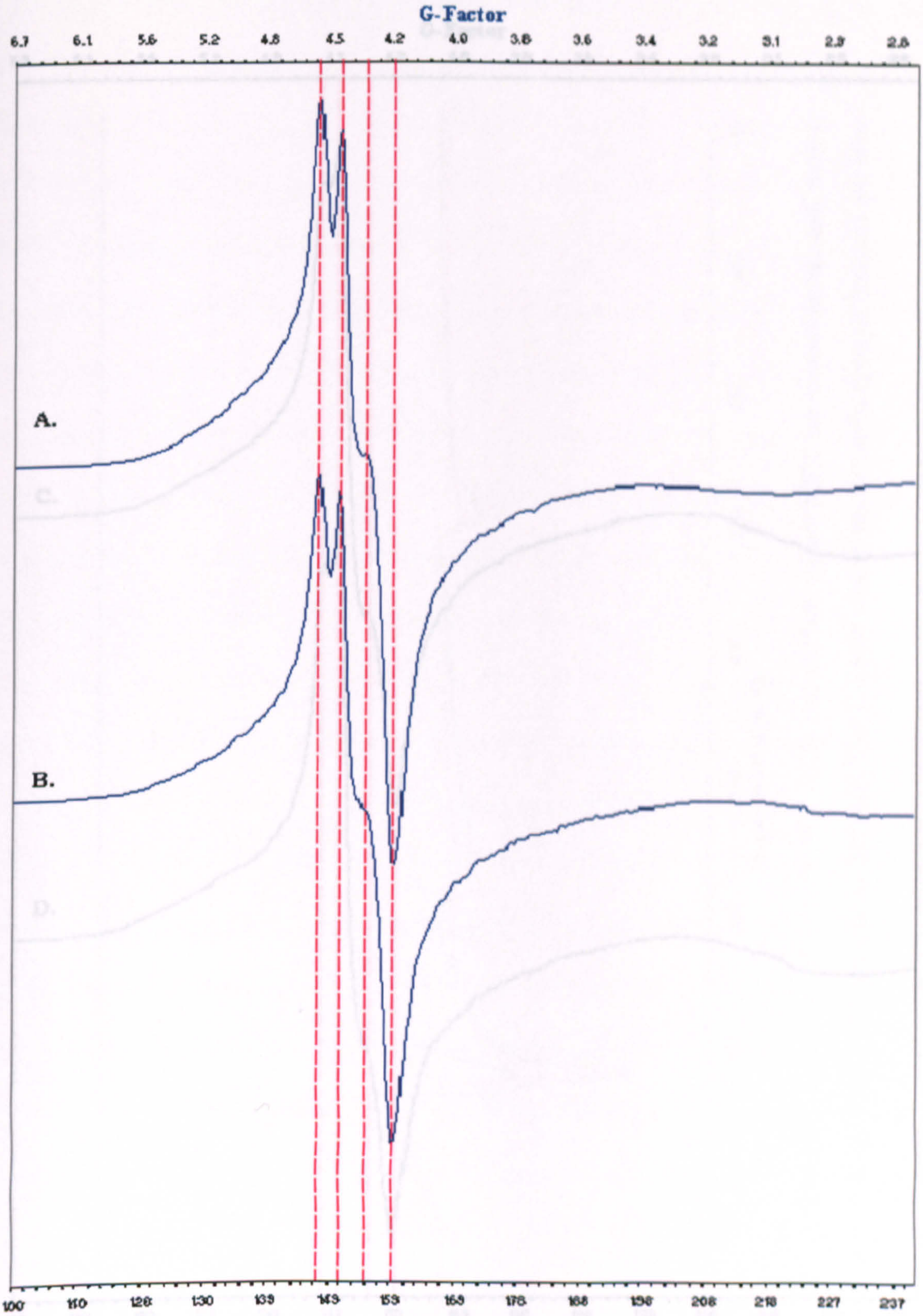


Figure 6.11. EPR Spectrum of diferric WT native (A) and recombinant (B) hSTf, scanning a magnetic field from 100 to 245 mT. The G-factor values for the main spectral features are also labelled.

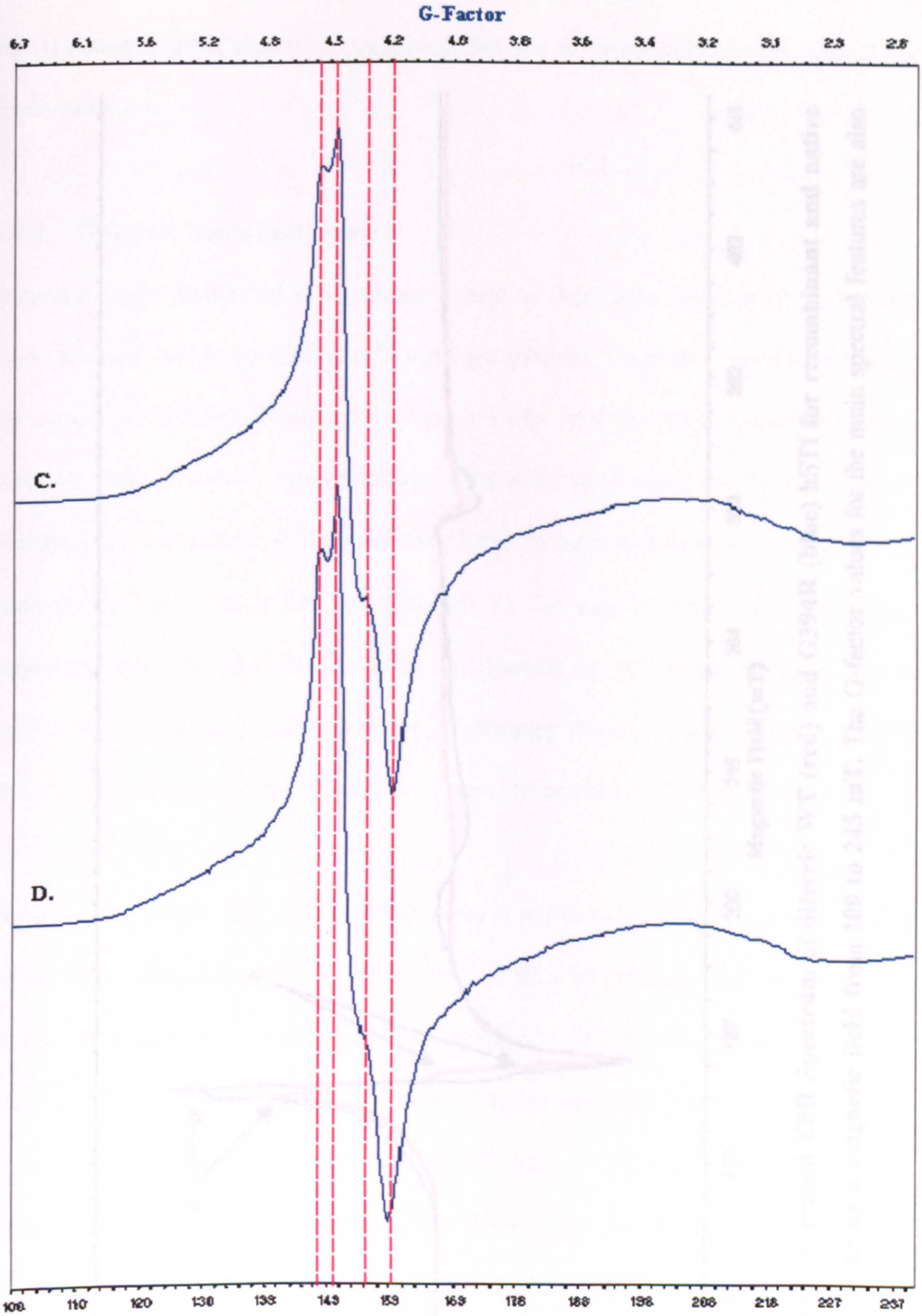


Figure 6.12. EPR Spectrum of diferric G394R native (C) and recombinant (D) hSTf, scanning a magnetic field from 100 to 245 mT. The G-factor values for the main spectral features are also labelled.

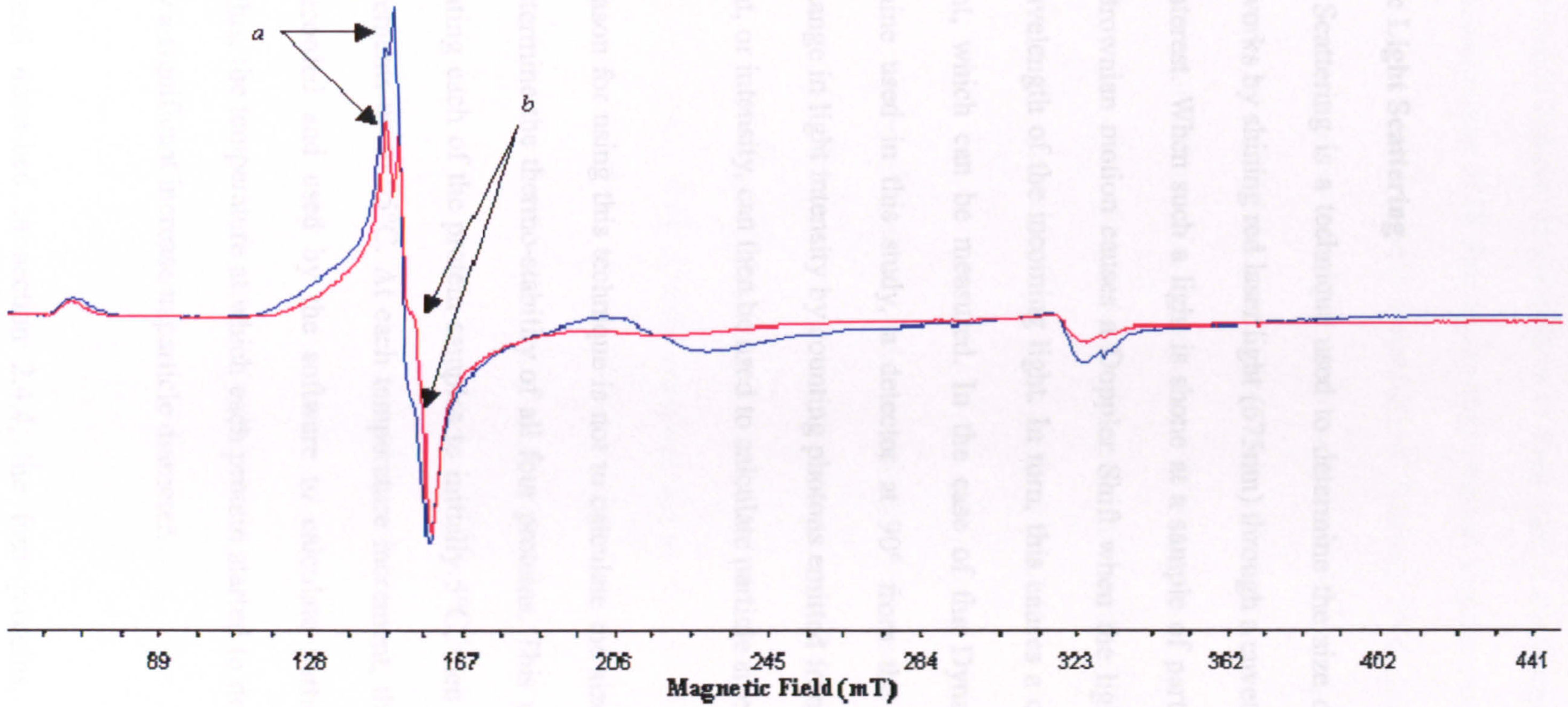


Figure 6.13. Averaged EPR Spectrum of diferric WT (red) and G394R (blue) hSTf for recombinant and native proteins, scanning a magnetic field from 100 to 245 mT. The G-factor values for the main spectral features are also labelled.

field in the range 180 to 362mT. However, these differences could be due to the signal strength rather than direct change difference in chemical behaviour around the ferric ions.

6.1.3. Dynamic Light Scattering

Dynamic Light Scattering is a technique used to determine the size of molecules. This technique works by shining red laser light (675nm) through a cuvette containing the sample of interest. When such a light is shone at a sample of particles such as proteins, their Brownian motion causes a Doppler Shift when the light hits them, changing the wavelength of the incoming light. In turn, this causes a change in the intensity of light, which can be measured. In the case of the DynaPro99 Light Scattering machine used in this study, a detector at 90° from the light source measures this change in light intensity by counting photons emitted from the sample. The photon count, or intensity, can then be used to calculate particle size.

However, the reason for using this technique is not to calculate the size of particles per se, but to determine the thermo-stability of all four proteins. This was achieved by gradually heating each of the protein samples to initially 5°C, then to 15°C and then in 5°C increments up to 95°C. At each temperature increment, the number of photons were recorded and used by the software to calculate particle (protein) diameter. From this, the temperature at which each protein started to denature could be determined by a significant increase in particle diameter.

Using the protocol described in section 2.4.4, the four proteins, native WT, recombinant WT, native G394R and recombinant G394R were analysed in both their

apo- and diferric forms. The results from this study (Figure 6.14) indicate that all four proteins are stable up to at least 55°C, in both their apo- and diferric forms, and that both recombinants behave very similarly to their native counterparts. Between 60-65°C, the average radius of all four apo-proteins almost simultaneously increases significantly in size, indicating protein denaturation. Towards 70°C, there is a slight increase in the average radius for the native and recombinant G394R diferric proteins, but between 80-85°C these proteins are seen to fully denature. Finally, the native and recombinant WT proteins denature at 85-90°C, indicating that these are the most heat-stable. It can therefore be concluded from this study that the binding of iron to hSTf stabilises its structure, and provides a higher resistance to heat.

These findings are consistent with the work published by Ikemoto and Ventura, 1979, who found through Differential Scanning Calorimetry that apo-hSTf denatured between 60-70°C and diferric hSTf denatured around 83°C. However, unlike the results obtained in our study, they actually observed that the apo-hSTf denatured in two steps (60.1°C and 70.7°C), as also described by Evans *et al*, 1982, who found apo-hSTf denatured at 64.4°C and then 74.4°C. Such two-step denaturation of apo-hSTf could be masked by the 5°C increments used in our study, which prevents the resolution of two separate denaturation events, which appear as a single event.

The fact that both recombinant and native G394R hSTf in their apo forms denature around 60-65°C is also consistent to previously published work. Evans *et al*, 1982 showed that native G394R hSTf denatures at 63.4°C and then 75.3°C in its apo form. This is only around 1°C less than the temperature they published for the WT protein.

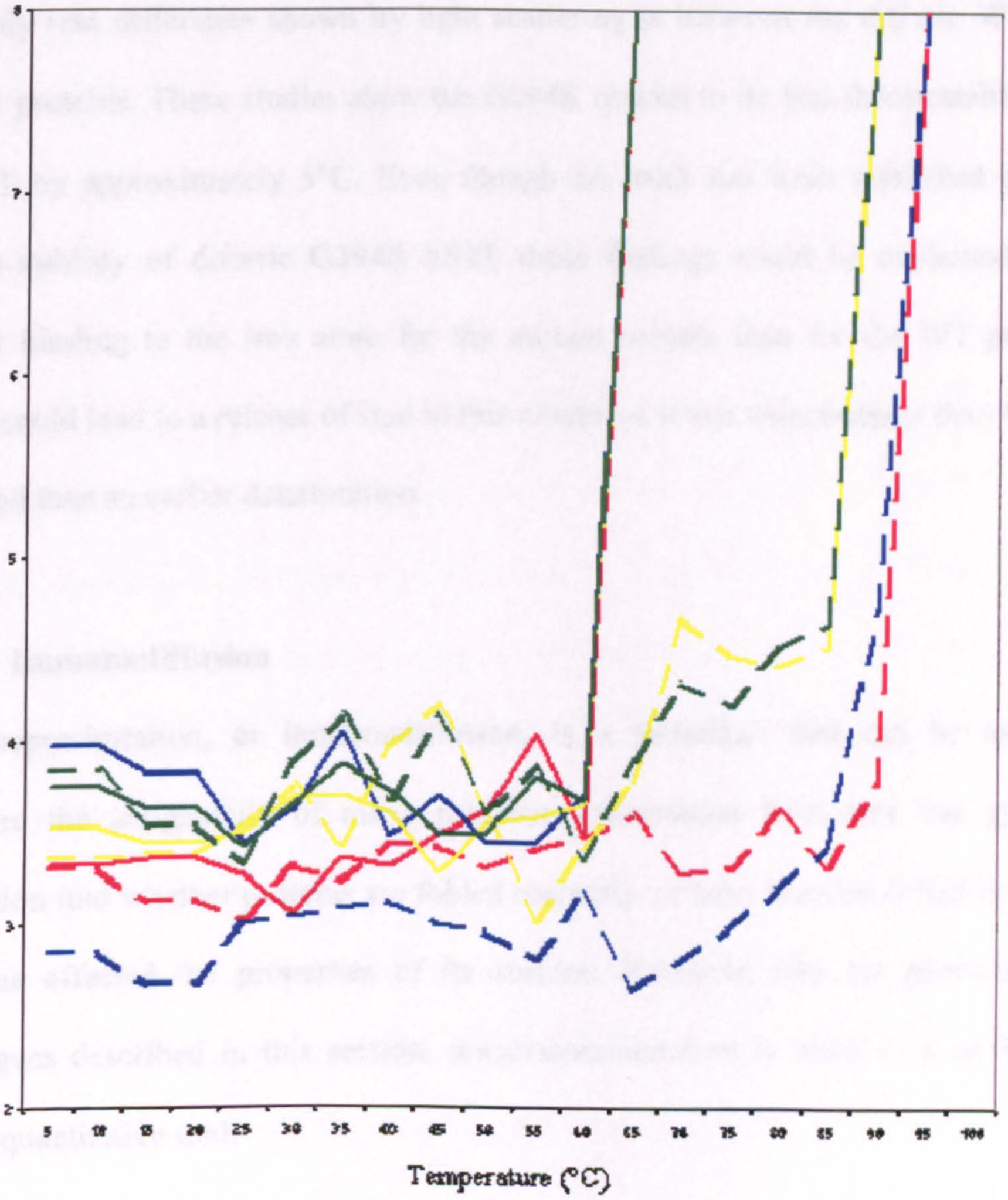


Figure 6.14. Dynamic Light Scattering data showing the thermostability of apo- and diferric hSTf. Native WT: red, recombinant WT: blue, native G394R: yellow and recombinant G394R: green. Apo samples are shown here as solid lines, whereas the diferric samples are shown here as dotted lines. Reading with a radius of ≥8 indicates complete denaturation.

These results therefore suggest that there is no significant difference in stability between the WT and G394R apo-proteins.

The only real difference shown by light scattering is between the diferric WT and G394R proteins. These studies show the G394R mutant to be less thermostable than the WT by approximately 5°C. Even though no work has been published on the thermo-stability of diferric G394R hSTf, these findings could be explained by a weaker binding to the iron atom for the mutant protein than for the WT protein, which could lead to a release of iron in this mutant at lower temperatures than for the WT, and thus an earlier denaturation.

6.1.4. Immunodiffusion

Immunoprecipitation, or immunodiffusion, is a technique that can be used to compare the antigenicity of many proteins. Information from this can give an indication into whether proteins are folded correctly, or have been modified in a way that has affected the properties of its surface. However, like the previous two techniques described in this section, immunoprecipitation is more of a qualitative than a quantitative tool.

This technique relies upon the diffusion of a multivalent antigen towards either a divalent (IgG) or multivalent (IgM) antibody in agar. If both molecules immunochemically react, large antibody-antigen complexes give rise to a molecular lattice that is too large to remain in solution, causing the formation of an immunoprecipitin line. In double immunodiffusion, when two antigens are diffused towards a single antibody, two immunoprecipitin lines can form if both antigens

react with the antibody. How both of these lines meet gives an indication of how immunochemically similar the two antigens are. If both lines meet, the two antigens are suggested to be immunochemically identical, whereas, if both lines cross, the two antigens are suggested to be immunochemically non-identically. When one line fuses and one does not, this suggests partial identity. The antigen that fuses shares all its determinants with the second one. However, all the determinants on the second one are not all shared with the first one (Wilson and Walker, 2000).

Using the protocol described in section 2.4.8, immunodiffusion slides were set up to compare the immunochemical similarity of all four proteins: native WT, native G394R, recombinant WT and recombinant G394R proteins in both their apo- and diferric- forms. All proteins were made either apo or diferric following the protocol described in section 2.3.5. As with the EPR study, both of the recombinants used were fully processed. The results from this study are shown in Figure 6.15.

Slides A, B and C are double diffusions of the four proteins in their apo form. Although the immuno-precipitin lines obtained are not that sharp, it can be seen that all lines fuse together, implying immunochemical identity for all four proteins. Slide C, D and E are also double diffusions of the four proteins, but in their diferric forms. Again, all lines fuse together infer immunochemical identity. Slide G and H compares the apo- and diferric proteins. The reason for repeating the immunodiffusion, with apo- and diferric proteins, is to see whether proteins show a conformational change between the two forms that could result in the antigenic determinants on their surface being altered. No difference could be seen between the two forms.

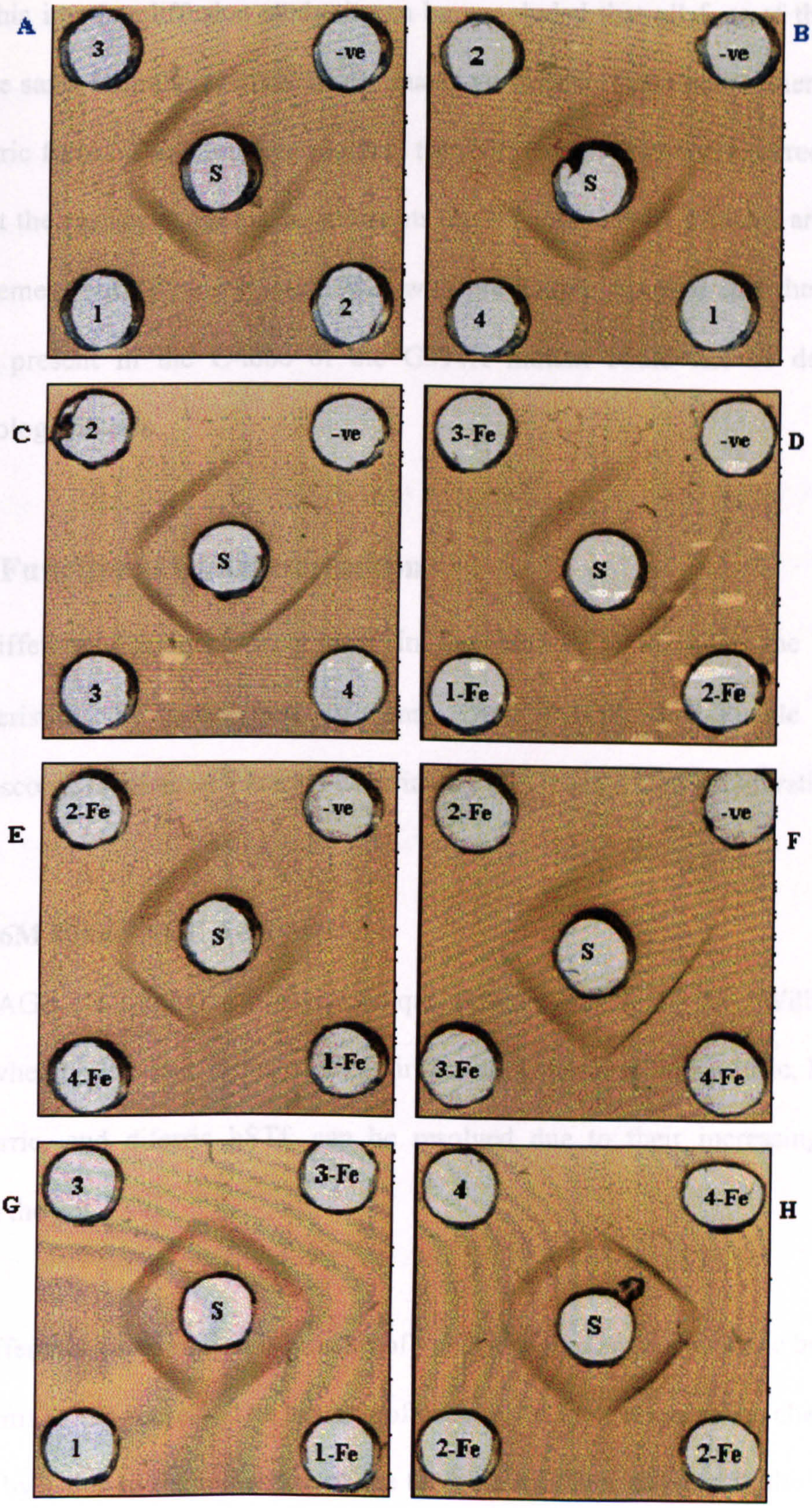


Figure 6.15. Immunodiffusion study of both native and recombinant, WT and G394R mutant hSTfs. S: Sheep Anti-hSTf serum, –ve: MilliQ Water, 1: Native WT hSTf, 2: Native G394R hSTf, 3: Recombinant WT hSTf, 4: Recombinant G394R hSTf and –Fe: Diferric.

From this immunodiffusion study, it can be concluded that all four of the proteins have the same immunochemical determinants on their surface under either their apo- or diferric forms. This therefore suggests that both recombinants are correctly folded and that the surface antigenic determinants on WT and G394R proteins are identical in agreement with Evans *et al*, (1982) who previously reported that the structural change present in the C-lobe of the G394R mutant could not be detected by immunological tests.

6.2. Functional Characterisation

Four different techniques were used in this section to compare the functional characteristics of these four proteins: Urea PAGE, UV-Visible Scanning Spectroscopy, Isothermal Titration Calorimetry (ITC) and a Cell Proliferation Assay.

6.2.1. 6M Urea PAGE Analysis

Urea PAGE electrophoresis is a technique (Makey and Seal 1976, Williams *et al* 1978) whereby the four forms of hSTf, iron-free, C-terminal monoferric, N-terminal monoferric, and diferric hSTf, can be resolved due to their increasing mobility through the gel.

This difference in migration is a result of variations in charge and shape between the four forms of the protein. The binding of each ferric ion decreases the charge on the protein by 1 due to the ionisation of the tyrosine residues, along with the binding of the bicarbonate anion, thus increasing the proteins mobility. The presence of 6M-urea causes the iron-free domains to unfold, which decreases the mobility of iron-free transferrin, and the two monoferric transferrins, relative to that of the iron-

saturated protein. Even though both monoferric forms are located in between the apo and diferric forms, Evans and Williams showed in 1980 that, due to the structure of transferrin, they both migrate slightly differently from each other, with the N-lobe monoferric having the highest mobility. The technique of 6M urea PAGE can therefore be used to study the binding of iron to and the release of iron from transferrin.

Using the protocol described in section 2.4.14 all six of the proteins in both their apo and diferric forms, including the fully and partially processed fractions of the recombinant proteins, were analysed by 6M Urea gel PAGE (Figure 6.16). In order to test these proteins for their ability to bind and release iron, all proteins were initially fully saturated with iron using the protocol described in section 2.3.5 and then a fraction of this diferric protein was then made apo using the protocol described in that same section.

Gel A (Figure 6.16) illustrates the different forms of the WT native and recombinant proteins, which imply the ability to bind and release iron for all three proteins. However, all three of the proteins migrate slightly differently from each other. Both of the recombinant proteins do not migrate as far as the native protein and fraction 1 does not migrate as far as fraction 2 of the recombinants. A probable reason for these differences could be the absence of the N-linked glycosylations on the recombinant proteins. Of the total serum hSTf, 80% is under tetrasialotransferrin form, carrying therefore 4 supplementary negative charges due to the four terminal sialic acid groups (Rosalki, 2004 and Jamieson *et al*, 1971). As a result, the net charge of recombinant transferrin molecules on 6M urea PAGE (pH 8.4), which do not have

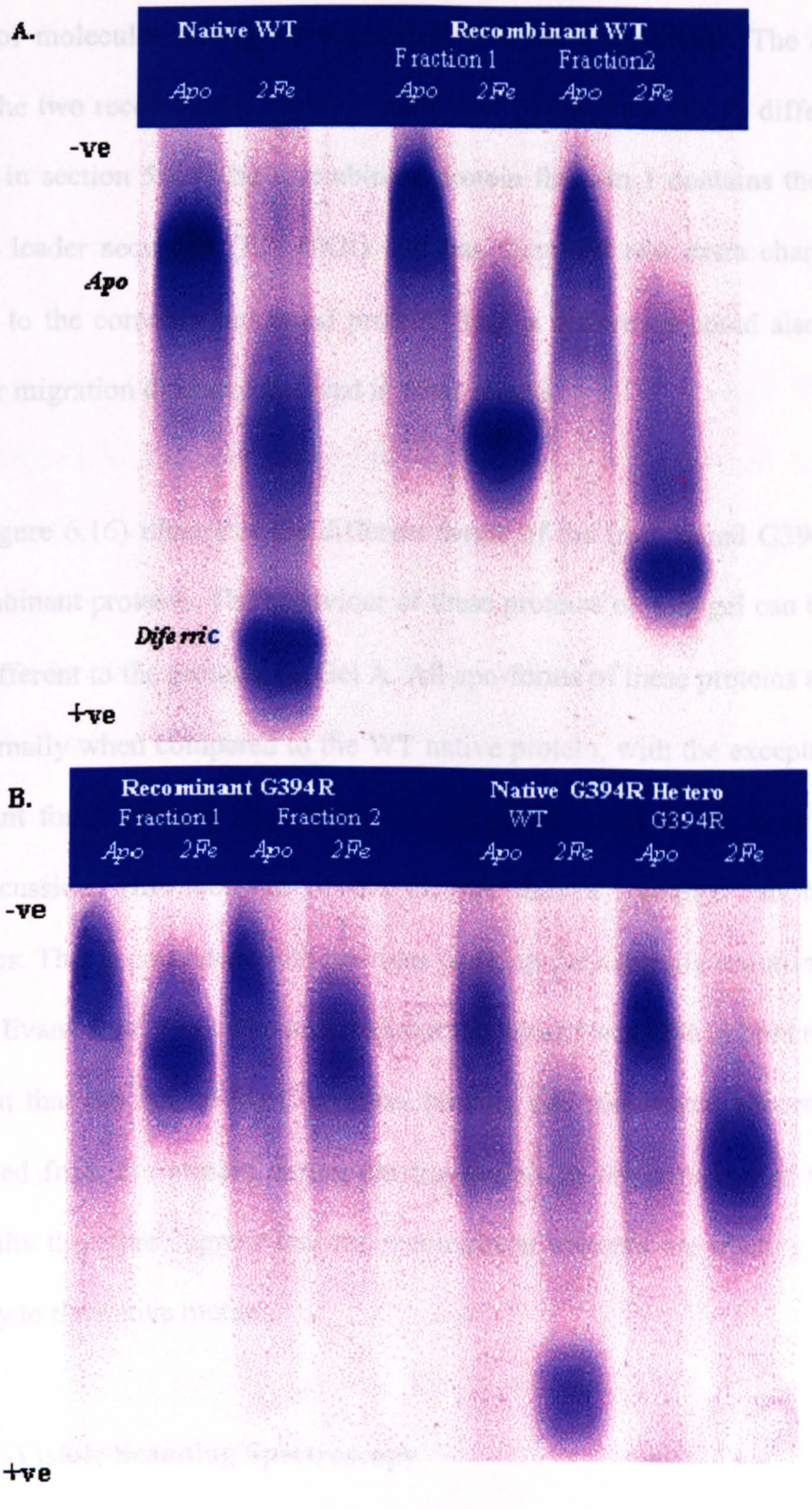


Figure 6.16. 6M Urea PAGE to study the iron-binding capabilities of native and recombinant, WT and G394R hSTfs. Gel A: Recombinant and native WT. Gel B: Recombinant and native G394R. Apo: Apo-transferrin. 2Fe: Diferric-transferrin.

any sialic acids, will be increased by four extra positive charges, reducing the mobility of molecules during electrophoresis towards the cathode. The difference between the two recombinant fractions could also be due to a charge difference. As discussed in section 5.1.5, the recombinant protein fraction 1 contains the partially processed leader sequence (RSLDKR) and has therefore two extra charges when compared to the correctly processed protein. Such a difference could also result in the shorter migration distance observed in these results.

Gel B (Figure 6.16) illustrates the different forms of the iron bound G394R native and recombinant proteins. The behaviour of these proteins on this gel can be seen to be very different to the proteins on Gel A. All apo-forms of these proteins are shown to run normally when compared to the WT native protein, with the exception of the recombinant forms running slightly behind the native forms as expected from the above discussion. However, all of the G394R diferric samples run more like monoferrics. This is consistent with previous work on the naturally occurring G394R mutant by Evans *et al*, 1982 who reported that the variant was able to bind two atoms of iron, but that the iron in the C-terminal binding site was bound abnormally and was released from the protein during electrophoresis in the presence of 6M-Urea. These results therefore suggest that the recombinant mutants are binding iron in a similar way to the native mutant.

6.2.2. UV-Visible Scanning Spectroscopy

UV-Visible scanning spectroscopy is a common technique used to study iron binding and release by transferrins. Lehrer (1969) reported that, following scanning of hSTf over the spectral range from 250 to 600nm, characteristic peaks in the region of

tyrosine and tryptophan absorption indicated perturbation of these chromophores by the binding of metal ions. More specifically, binding of Fe^{3+} and Cu^{2+} yielded peaks at around 470nm and 440nm respectively. Therefore, it is possible to obtain a characteristic fingerprint of ironbound transferrin and also to determine the iron-binding capacity of the protein by titration of iron until saturation occurs.

Using the protocol described in section 2.4.15 each of the four proteins: WT native, WT recombinant, G394R recombinant and G394R native were scanned from 350 to 600nm, repeating the scan for each titration of FeNTA (Figure 6.17 and 6.18). Proteins were made apo using the protocol described in section 2.3.5. These spectra show a characteristic peak in absorbance around the 420-480nm region for all four proteins, with each titration causing an increase in absorbance around this range until protein saturation occurs. This stepped increase in absorbance, following FeNTA titration, is indicative of the protein's ability to bind iron.

When comparing the spectral data collected from the iron-saturated proteins (Figure 6.19), a number of differences can identified between the proteins. One major difference was the final absorbance attained, which can be explained by a slightly different concentration for each sample. Measuring the absorbance of each protein in its apo-form at 280nm, and using the equations in section 2.4.15, its protein concentration could be calculated. It was then possible to calculate the theoretical absorbance at 470nm using the same equations. This theoretical absorbance value can give an indication of how much of the protein is binding iron. If there is 100% binding, the observed absorbance at 470nm should be close to the calculated theoretical value (approximately 19 times less than the absorbance of the apo protein

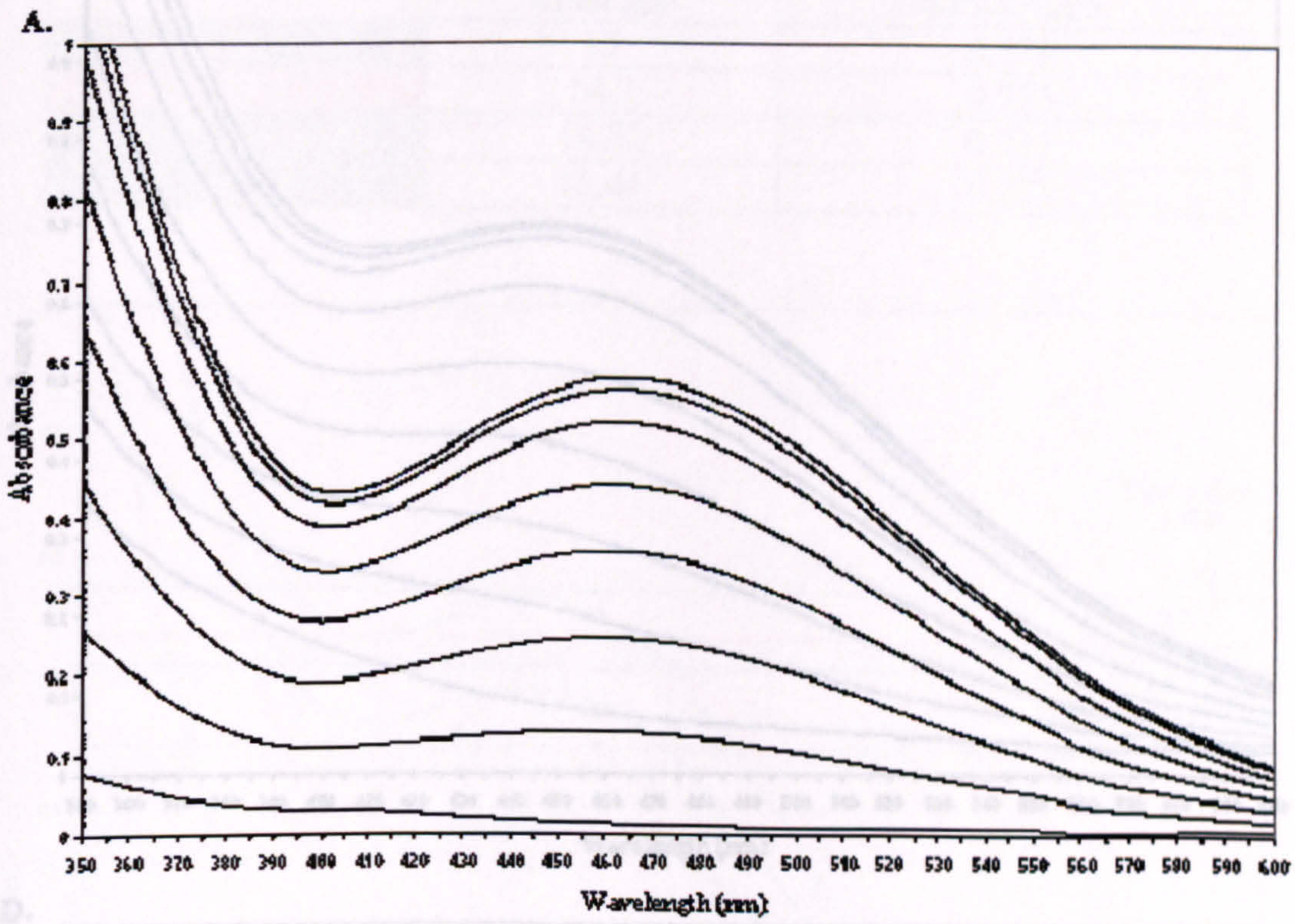


Figure 6.17. Scanning UV-visible spectra (350-600nm) of iron-free WT native (A) and WT recombinant (B) hSTf following titration with 5µl aliquots of 0.01M FeNTA until proteins become iron saturated.

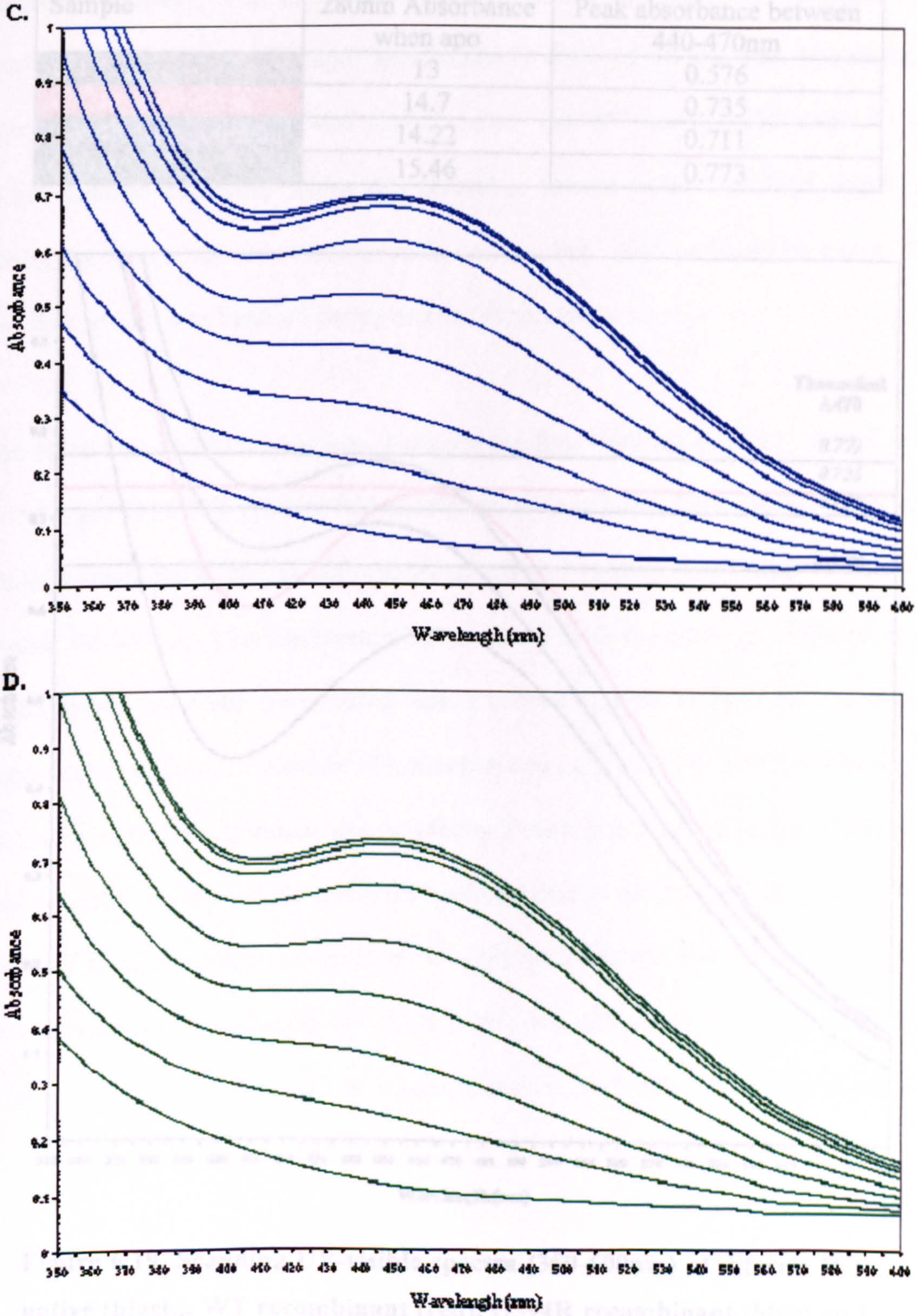


Figure 6.18. Scanning UV-visible spectra (350-600nm) of iron-free G394R native (A) and G394R recombinant (B) hSTf following titration with 5µl aliquots of 0.01M FeNTA until proteins become iron saturated.

Sample	280nm Absorbance when apo	Peak absorbance between 440-470nm
WT Native	13	0.576
WT Recombinant	14.7	0.735
G394R Recombinant	14.22	0.711
G394R Native	15.46	0.773

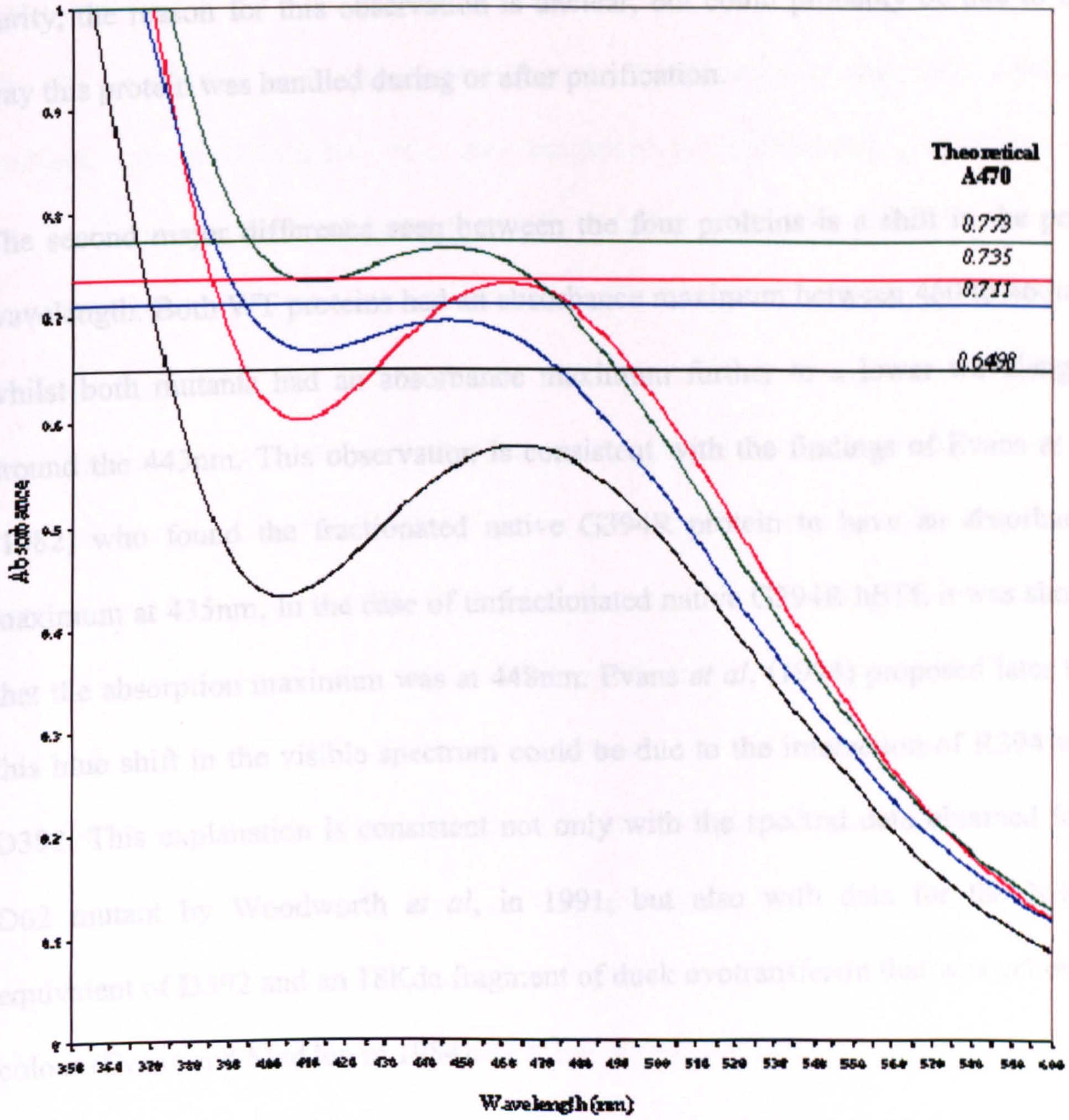


Figure 6.19. Scanning UV-visible spectra (350-600nm) of diferric WT native (black), WT recombinant (red), G394R recombinant (blue) and G394R native (green) hSTf. The actual absorbance of apo protein at 280nm and peak absorbance between 440-470nm when iron saturated is stated in the above table, and the theoretical 470nm maximum for all four proteins, based on their 280nm absorbance, is highlighted on the chart.

at 280nm). For each of the four proteins, the theoretical absorbance at 470nm was therefore compared with the actual peak absorbance between 440-470nm. These results showed that on average, approximately 96% of all proteins bound iron. However, this data was skewed slightly, since the WT native hSTf only had an 88.6% binding capability. Since this protein is known to be of a high degree of purity, the reason for this observation is unclear, but could probably be due to the way this protein was handled during or after purification.

The second major difference seen between the four proteins is a shift in the peak wavelength. Both WT proteins had an absorbance maximum between 460 to 465nm, whilst both mutants had an absorbance maximum further to a lower wavelength, around the 443nm. This observation is consistent with the findings of Evans *et al*, (1982) who found the fractionated native G394R protein to have an absorbance maximum at 435nm. In the case of unfractionated native G394R hSTf, it was shown that the absorption maximum was at 448nm. Evans *et al*, (1994) proposed later that this blue shift in the visible spectrum could be due to the interaction of R394 with D392. This explanation is consistent not only with the spectral data obtained for a D62 mutant by Woodworth *et al*, in 1991, but also with data for the N-lobe equivalent of D392 and an 18Kda fragment of duck ovotransferrin that was yellow in colour (Evans and Madden in 1984).

6.2.3. Isothermal Titration Calorimetry

Isothermal Titration Calorimetry (ITC) is a thermodynamic technique for monitoring any chemical reaction initiated by the addition of a binding component. In this system, when two molecules interact, heat is either generated or absorbed, depending

on whether the interaction is exothermic or endothermic. By accurately measuring this change in heat, ITC can allow accurate determination of the association constants (K_a), reaction stoichiometry (n), free energy (ΔG), enthalpy (ΔH) and entropy (ΔS). Therefore ITC can provide a complete thermodynamic profile of molecule interactions.

The ITC machines works by maintaining the temperature of two cells using a feedback system. These two cells are designated the sample cell, containing the macromolecular solution of interest, and the reference cell, containing either buffer or water. During the injection of titrant into the sample cell, heat is taken up or evolved depending on whether the macromolecular association reaction is endothermic or exothermic. In an exothermic reaction, the temperature of the sample cell will increase and the feedback power will be deactivated to maintain equal temperatures between the cells. For an endothermic reaction, the opposite will take place. That is, the feedback power will be turned on (Pierce *et al*, 1999). During an ITC experiment, the direct observable measurement is the time-dependent input required to maintain equal temperatures in the sample and reference cells (μcal/s). As successive amounts of the ligand are titrated into the cell, the quantity of heat absorbed or released is directly proportional to the amount of binding occurring with time. As the system reaches saturation, the signal diminishes until only heats of dilution are observed. A binding curve is then obtained from a plot of the 'heats' from each injection against the ratio of ligand and binding partner in the cell. From this curve, thermodynamic data for the molecular interactions and stoichiometry can be calculated.

Before analysing of the four proteins, apo preparations of native WT, native G394R, recombinant WT and recombinant G394R were made using the protocol described in section 2.3.5. Using the protocol described in section 2.4.9, all four hSTf proteins were then titrated separately with FeNTA in a VP-ITC (MicroCal, USA) calorimeter and the energy required to stabilise the temperature of the two cells, following each titration, recorded. The results from each titration are shown in Figure 6.20, 6.21, 6.22 and 6.23, for each of the four proteins, respectively. Each of these figures shows the differential power signal recorded during each experiment (μcal/sec versus time) (top) and a sigmoidal plot of kcal/mole versus molar ratio (below). This sigmoidal plot data was calculated from the integration of each differential power signal peak. The line of 'best fit' was generated automatically using the ORIGIN software, with slight manual adjustments to take into account experimental error.

The four figures show two major characteristics of these reactions. Firstly, the differential power signal indicates that iron binding by each protein is an exothermic event, since the ITC machine reduces its power input into the cell following each titration. This is indicative of the machine attempting to compensate for heat generated by the reaction. These results are consistent with published data for ovotransferrin (Mason *et al*, 1996). It can also be concluded from these figures that more than one binding event is occurring since the molar ratio at the point of inflection of the sigmoidal curve is greater than 1.5. A theoretically value of 2, which should have been obtained, is in fact rarely achieved due to experimental error and data interpretation difficulties. In reality, the number of binding events is generally calculated by rounding-up the molar ratio at this point.

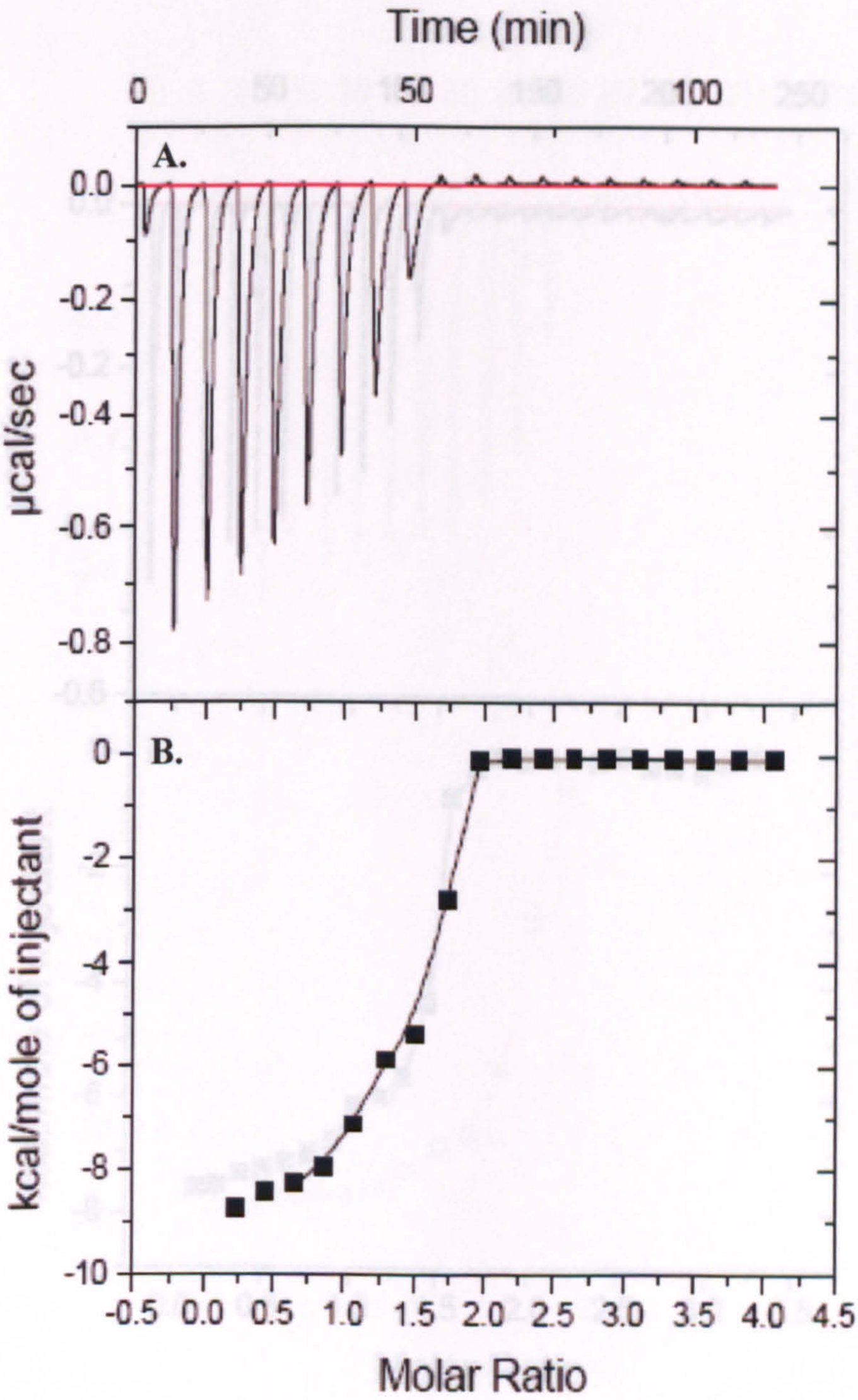


Figure 6.20. ITC data of native WT hSTf being titrated with FeNTA up to and beyond saturation. A: Differential power signals during titration of hSTf with FeNTA. B: Integrated differential power signal data plotted as a sigmoidal curve.

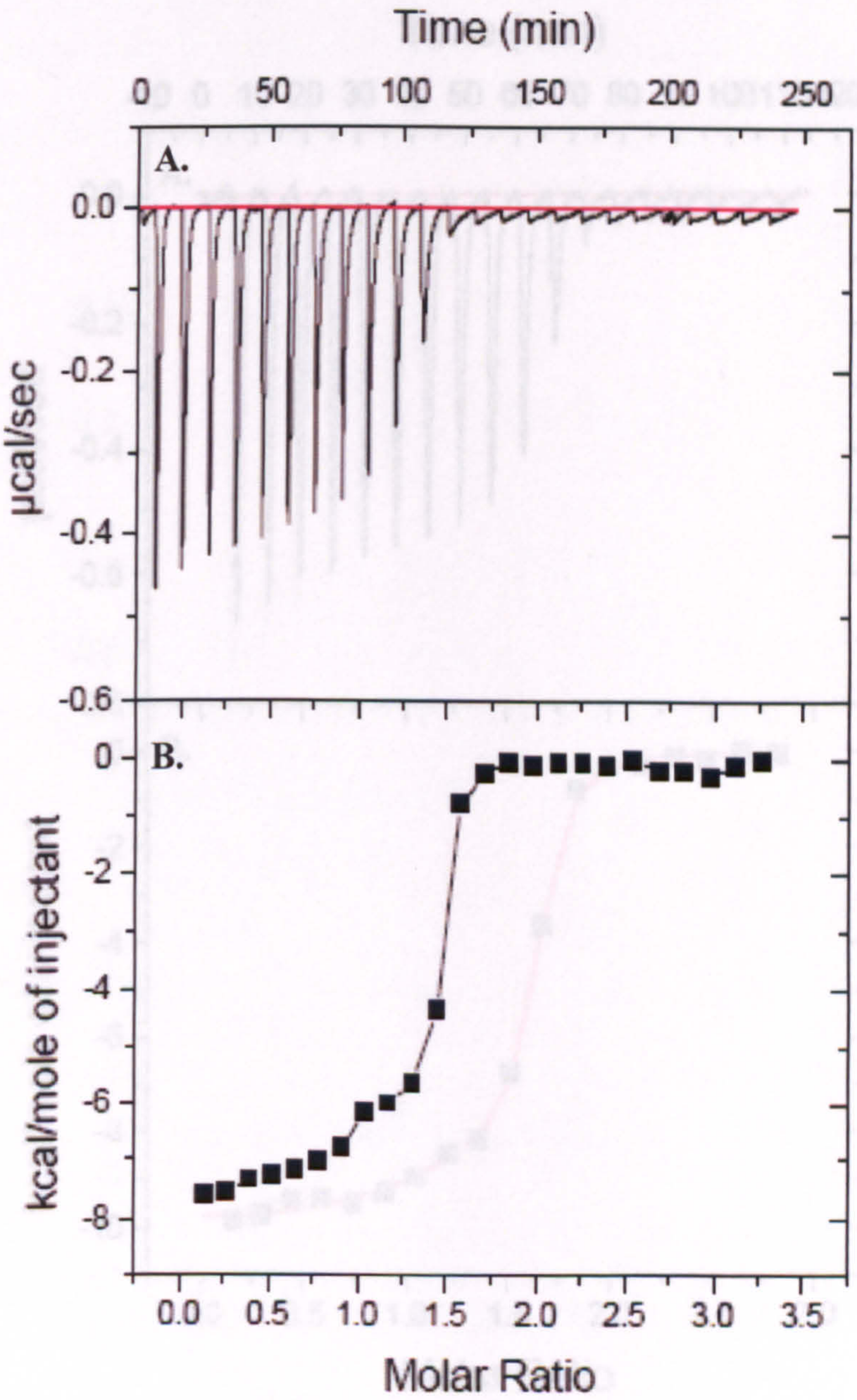


Figure 6.21. ITC data of recombinant WT hSTf being titrated with FeNTA up to and beyond saturation. A: Differential power signals during titration of hSTf with FeNTA. B: Integrated differential power signal data plotted as a sigmoidal curve.

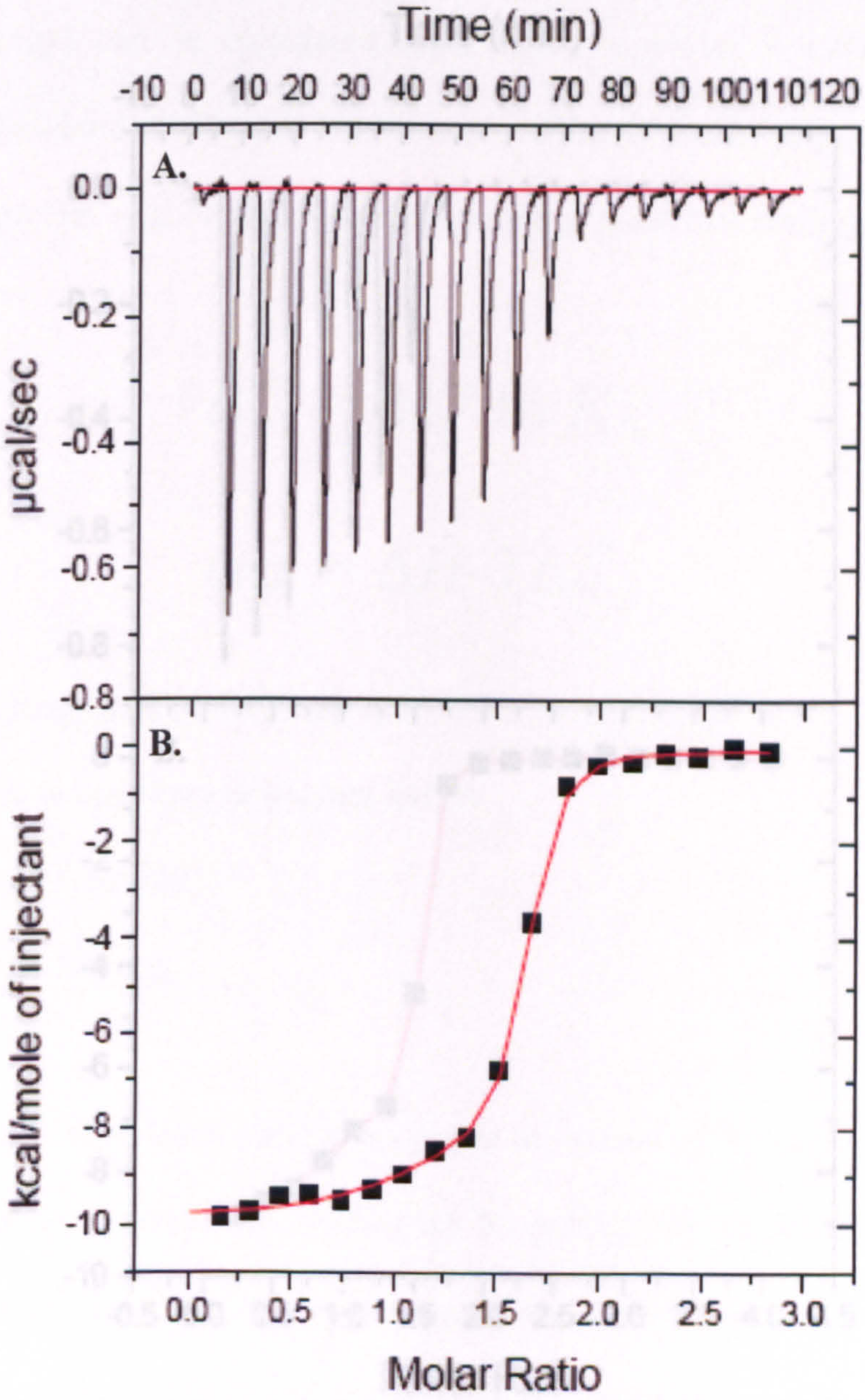


Figure 6.22. ITC data of native G394R hSTf being titrated with FeNTA up to and beyond saturation. A: Differential power signals during titration of hSTf with FeNTA. B: Integrated differential power signal data plotted as a sigmoidal curve.

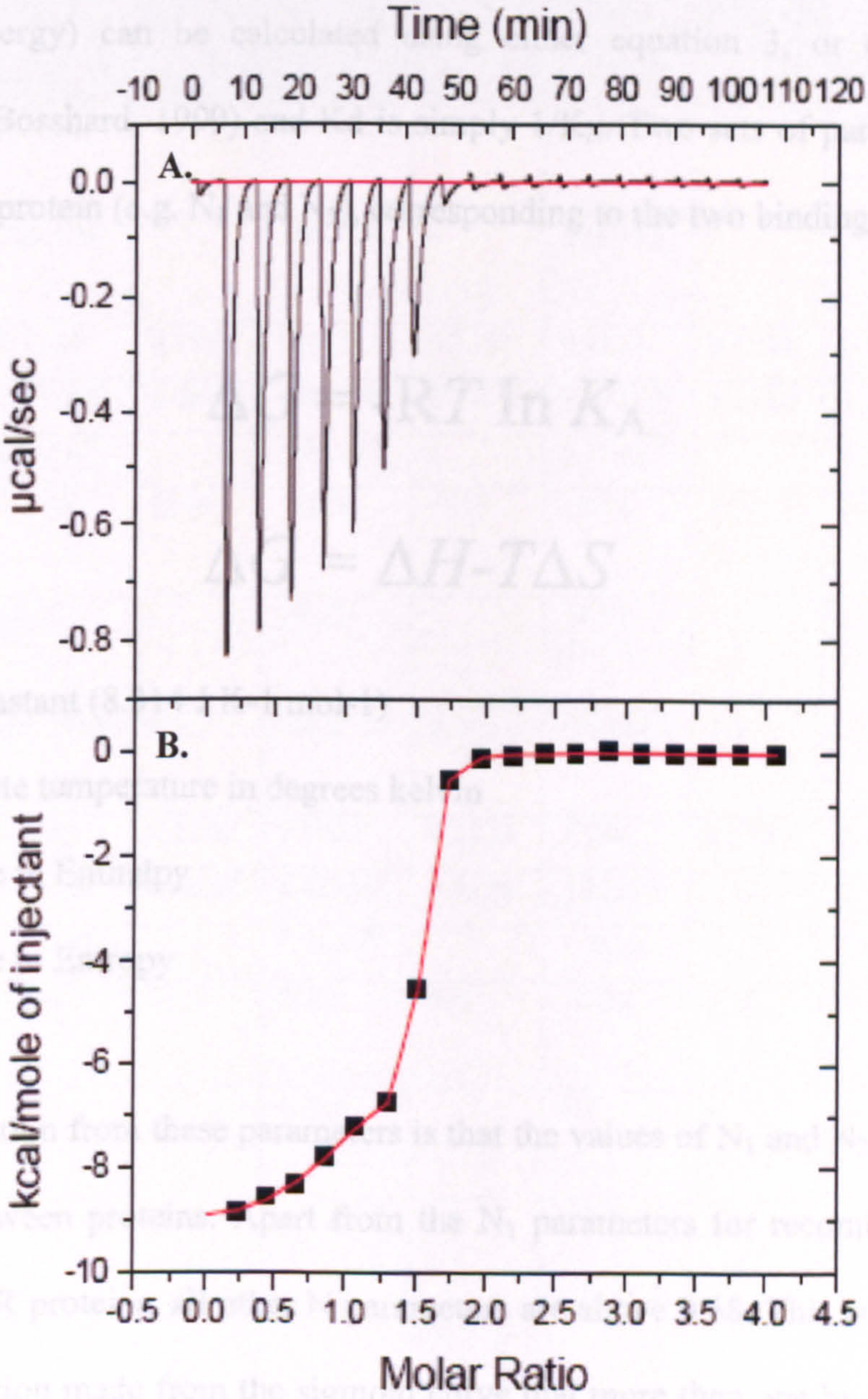


Figure 6.23. ITC data of recombinant G394R hSTf being titrated with FeNTA up to and beyond saturation. A: Differential power signals during titration of hSTf with FeNTA. B: Integrated differential power signal data plotted as a sigmoidal curve.

From these sigmoidal curves, the ORIGIN software automatically calculates the thermodynamic parameters for each reaction. A comparison of these parameters is shown in Table 6.1 along with ΔG parameters and dissociation constants (Kd). ΔG (Gibbs free energy) can be calculated using either equation 3, or equation 4) (Jelesarov and Bosshard, 1999) and Kd is simply $1/K_A$. Two sets of parameters are shown for each protein (e.g. N_1 and N_2), corresponding to the two binding sites.

$$\Delta G = -RT \ln K_A \qquad [3]$$

Or...

$$\Delta G = \Delta H - T\Delta S \qquad [4]$$

-R = gas constant (8.314 J K-1 mol-1)

-T = Absolute temperature in degrees kelvin

-ΔH = Change in Enthalpy

-ΔS = Change in Entropy

The first observation from these parameters is that the values of N_1 and N_2 vary quite substantially between proteins. Apart from the N_1 parameters for recombinant WT and native G394R proteins, all other N parameters are above 0.68. This is consistent with the observation made from the sigmoid curve that more than one binding event is occurring. The reason for the N_1 parameters for recombinant WT and native G394R proteins being lower than 0.5 is most likely due to experimental error and difficulties in fitting the sigmoid curve to the data points. Taking into account the margins of error stated in Table 6.1, N_1+N_2 for all four proteins can be calculated to be close to 2.

hSTf / Paramaters	Native WT	Error (+/-)	Recomb WT	Error (+/-)	Native G394R	Error (+/-)	Recomb G394R	Error (+/-)
Temperature (°C) (T)	25.0161	-	25.0145	-	25.0125	-	25.0086	-
Temperature (Kelvin)	298.176	-	298.174	-	298.172	-	298.168	-
N_1	0.7124	0.1993	0.4421	0.1180	0.4833	0.1358	0.6884	0.0243
K_1 (M^{-1})	$7.95x10^6$	$6.0x10^6$	$1.99x10^7$	$2.5x10^7$	$5.97x10^6$	$3.2x10^6$	$1.56x10^7$	$1.4x10^6$
Kd (M^{-1})	$1.3x10^{-7}$		$5.0x10^{-8}$		$1.7x10^{-7}$		$6.4x10^{-8}$	
ΔH_1 (J mol^{-1})	-4340	3106	-4229	1815	-5914	1187	-6407	198.6
ΔS_1 (J K^{-1} Mol^{-1})	17.02	-	19.22	-	11.17	-	11.43	-
ΔG_1 (kJ mol^{-1})	-39.3908	1.4	-41.6717	2.0	-38.6801	1.1	-41.0713	0.2
N_2	0.9185	0.2174	0.9822	0.1236	1.234	0.1449	0.7708	0.0259
K_2 (M^{-1})	$4.16x10^7$	$4.1x10^7$	$8.42x10^7$	$9.3x10^7$	$5.82x10^6$	$2.5x10^7$	$1.42x10^7$	$3.0x10^6$
Kd (M^{-1})	$2.4x10^{-8}$		$1.2x10^{-8}$		$1.7x10^{-7}$		$7.0x10^{-8}$	
ΔH_2 (J mol^{-1})	-7674	1275	-8051	468	-7828	179	-9131	103.9
ΔS_2 (J K^{-1} Mol^{-1})	9.124	-	9.259	-	4.7	-	6.687	-
ΔG_2 (kJ mol^{-1})	-43.4954	1.7	-45.2410	1.8	-38.6146	1.1	-40.8412	0.5

Table 6.1. Calculated parameters for ITC titration of FeNTA into native WT-, recombinant WT-, native G394R- and recombinant G394R hSTf. All parameters and margin of error, apart from ΔG data, were calculated automatically by ORIGIN (MicroCal) software.

From these experiments, association constants (K_A) and dissociation constants (Kd) can be calculated. However, for very tight binding reactions, when K_A is higher than 10^9 M^{-1}, even the most sensitive of ITC instruments cannot accurately calculate binding constants (Jelesarov and Bosshard, 1999). Therefore, these results are more for comparative purposes rather than accurate measurement of binding constants. Even so, the results obtained in these studies for both the WT recombinant and native proteins are very similar to the results published by Lin *et al*, 1993. The small differences between the two studies are probably due to slight differences in experimental temperature (25°C vs 27°C) and buffers.

When both the K_A and Kd for both WT proteins, and for each binding event are compared, it can be seen that there is on average a 10 fold difference in binding affinity between the two binding sites. This is again consistent with the results published by Lin *et al*, 1993 and can possibly be explained by the C-lobe having a stronger affinity for iron than the N-lobe (Aisen *et al*, 1978). Therefore, in the case of these WT proteins, K_1 is likely to correspond to the N-lobe, whilst K_2 probably corresponds to the C-lobe. In the mutant proteins however, even though the values for K_1 are similar to the K_1 values for the WT proteins, their K_2 values are between 6 to 7 times lower than the WT proteins K_2 values. In fact, the K_2 values for the mutant proteins are actually lower than their K_1 values. These findings are consistent with the work published by Evans *et al*, 1982, in that the G394R mutation seems to have reduced the iron binding affinity of the C-lobe. There also seems to be a slight difference between the native and recombinant proteins. For both the WT and mutant proteins, the recombinant proteins are shown here to have on average a 2.4x stronger iron binding affinity.

When the ΔH values for each protein and for both binding sites are compared, it can be seen that all iron-binding events are exothermic. However, the amount of energy produced from each binding event varies. ΔH_1 for the WT proteins is shown to be on average 2-fold higher than ΔH_2 and 1.4-fold greater than ΔH_2 for the mutant proteins. This is again consistent with Lin *et al*, 1993, and not with the observations of Binford and Foster, 1974, who suggest that the binding enthalpies for the N- and C-lobes are the same. Therefore, since both lobes are different, this suggests that they are both thermodynamically and kinetically different. When comparing the mutant to the WT proteins, the G394R mutation is shown to decrease ΔH for both lobes by approximately 1.2-fold. Interestingly this decrease is not just in the C-lobe, where the mutation is present, but actually can be seen in both lobes, which suggests that that both lobes are probably interconnected and influenced by each other.

Since enthalpy (ΔH), expressed as J $mole^{-1}$, can be related directly to the amount of energy required to break one mole of bonds between pairs of atoms, it can be concluded that a negative ΔH indicates the net increase in atomic interactions. From this, it can then be postulated that more molecular interactions are occurring in the G394R protein upon binding iron.

The values for entropy (ΔS) also varies between proteins and between sites. These studies show approximately a 2-fold decrease in ΔS between the N-lobe and C-lobe for all four proteins. This suggests that more disorder is occurring in the N-lobe during iron binding, compared to the C-lobe. Another way of looking at this is the N-lobe could be seen to be more flexible than the C-lobe on binding iron. A possible reason for this could be due to the C-lobe having a tighter binding affinity for iron

(section 1.3.4), therefore restricting movement of the two domains. Another reason could be the larger number of atomic interactions in this lobe on iron binding, shown by a lower ΔH. It is known however, that the N-lobe is more stable than the C-lobe to themal-denaturation (Donovan and Ross, 1976) and during electrophoresis on a 6M-urea gel (Evans *et al,* 1980), probably due to it having a higher number of disulphide bonds. Therefore, the stability of the N-lobe is probably more to do with the folding of the individual domains, rather than the flexibility of both domains together. It can also be seen by this study that the G394R mutation has resulted in a 1.6 fold decrease in ΔS in both lobes. Since a decrease in ΔS is indicative of an increase in molecular rigidity, it can therefore be interpreted as the G394R mutation restricting the movement of both domain in the C-lobe, but also reducing movement in the N-lobe.

Finally, the last parameter calculated from these ITC studies is ΔG. Since ΔH is negative for all binding events and ΔS is positive, ΔG will always be negative at any temperature of the reaction. Such a reaction will be spontaneous at all temperatures (Sutton *et al*, 2000). Since the WT proteins have a lower ΔG than the G394R mutant proteins, this suggests that the WT proteins have a greater ability to bind iron. This data also shows that for the WT proteins, the C-lobe has a greater tendency to bind iron than the N-lobe. However, for the mutant protein, this is reversed.

6.2.4. Cell Proliferation Assay

The abilities of the recombinant proteins to bind and release iron has been established in this study, but their ability to deliver iron to cells through receptor-mediated endocytosis has not yet been demonstrated. Therefore, the aim of this

section was to test the recombinant and native proteins in their ability to fulfil such a function. However, due to the lack of native G394R hSTf, this protein was not included in this part of the study and the only proteins to be investigated were the native WT hSTf, recombinant WT hSTf and the recombinant G394R hSTf.

Since iron is essential for cellular proliferation, its availability has a direct effect on growth rate (Trowbridge and Shackelford, 1986). Therefore, by using these three different hSTfs as the main source of iron in cell culture, their efficiency to deliver iron to the cells by receptor-mediated endocytosis can be determined by measuring levels of cell proliferation. In order to measure these levels of cell proliferation, a BrdU cell proliferation assay (Roche Applied Sciences) was used. This assay is based on the measure of DNA replication in cell proliferation before the cell divides into two daughter cells. By adding a labelled DNA precursor such as 5-Bromo-2'-deoxyuridine (BrdU) to the cell culture during cellular proliferation, cells take it up and incorporate it into the newly synthesised DNA. Therefore, as cellular proliferation increases, so will the amount of DNA and so will the amount of BrdU incorporated. This incorporated BrdU can then be detected by a quantitative cellular enzyme linked immunoassay using monoclonal antibodies directed against BrdU.

Using the protocol described in section 2.3.5, diferric preparations of the three different proteins were made together with an apo preparation of the native WT hSTf to be used as a negative control. These four proteins, together with a control of foetal bovine serum (FBS), were tested individually in collaboration with Dr Heinz Zoller (Department of Medicine, University of Cambridge), using the protocol described in section 2.4.2, for their ability to stimulate cell proliferation in two different cell lines.

The first cell line were Chinese hamster ovary (CHO) (TRVb-1) cells that lack both endogenous transferrin receptors, but since being transfected with TfR1 cDNA, stably expresses only TfR1 (Carlson *et al*, 2005). This allows the study to focus on protein interactions with only one receptor. The second cell line used was a human hepatoma cell (HepG2). These cells were used as a control as they express both human receptors for transferrin and human asialoglycoprotein receptor (Popovic and Templeton, 2004). Each of these studies were repeated five times, with each data set then averaged.

Figure 6.24 shows that all three diferric proteins significantly stimulate TRVb-1 and HepG2 cell proliferation more than apo-hSTf indicating that both of the recombinant proteins and the native protein are able to deliver iron to cells. However, their efficiency is shown to vary slightly with the most efficient being the native WT, followed by the recombinant WT and then the recombinant mutant. This result is consistent with published data from Evans *et al*, 1994, where they suggest that native G394R hSTf has a reduced affinity for the transferrin receptor. Alternatively, this difference may not be significant as the error bars for each of the three proteins are at similar levels. Therefore, no major difference can be seen from these results that show that all three proteins are able to deliver iron to TRVb-1 and HepG2 cells through receptor-mediated endocytosis. The ability of the recombinant protein to deliver iron to cells is important, since a pathogen free replacement of native transferrin in cell line supplements is advantageous.

It is worthy of notice that the proliferation stimulation observed with FBS on HepG2 cells is not observed with TRVb-1 cells. The effect of FBS on TRVb-1 proliferation

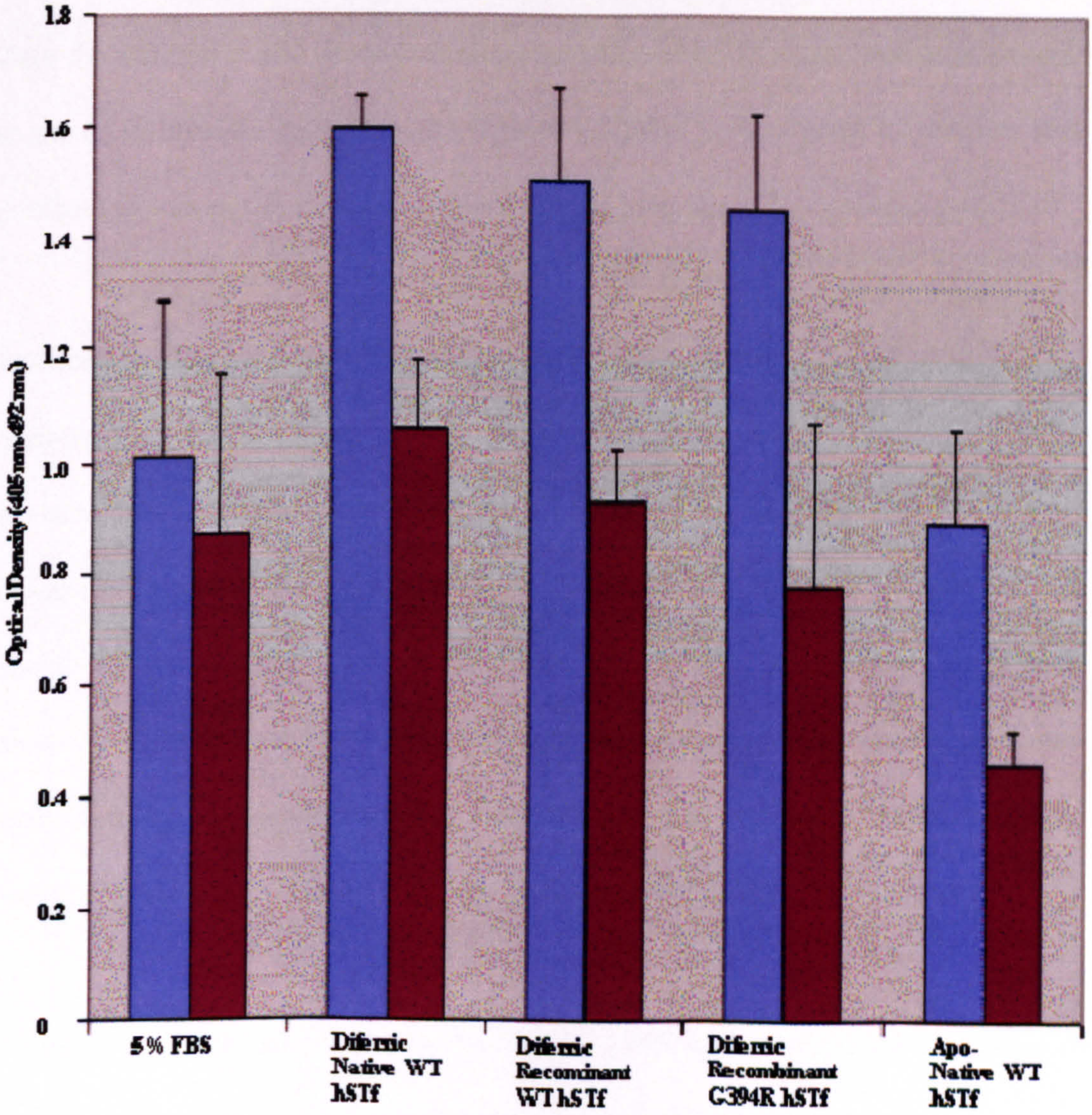


Figure 6.24. A BrdU assay to show the effect of using native and recombinant hSTf, including recombinant G394R hSTf, on the cellular proliferation of TRVb-1 (blue) and HepG2 (brown). The data shown is an average of five data sets.

is similar to that of apo-hSTf, suggesting that HepG2 cells have a greater ability than TRVb-1 cells in taking up iron and other essential nutrients from FBS. One possible reason for this could be due to the HepG2 human cells having a closer similarity to bovine cells than hamster (CHO) TRVb-1 cells.

7. Discussion

Amongst the Transferrin family, hSTf is now one of the best-characterised proteins. However, the mechanism by which it binds and releases iron, together with how it interacts with its receptors, is still poorly understood. Therefore, in order to further elucidate the structural and functional characteristics of hSTf, these studies focussed on the use of techniques such as computational analysis, recombinant expression and biophysical / biochemical analysis to attempt to address these fundamental questions.

As discussed in Chapter 4, two different cell types were explored for the creation of a recombinant expression system able to provide routinely the high amount of hSTf necessary to study its structure and function. The first to be explored was *E.coli*, due to the ease with which it can be manipulated and grown, and also the plethora of expression vectors available for use in such a system. The second type of system explored was *S.cerevisiae*. Even though these are not as easy to manipulate, as *E.coli* and only few expression vectors are available for use with such cells, a highly developed expression system provided by Delta Biotechnology Ltd was available.

The expression of the full-length (De Smit *et al*, 1995) and separate lobes (Steinlein and Ikeda, 1993) of hSTf has been reported in *E.coli*. However, this published work suggests that the expressed hSTf forms inclusion bodies and that protein refolding and purification of functional protein was difficult. The formation of such inclusion bodies probably occurs due to incorrect protein folding and is either due to the cells attempting to express hSTf at too high a level, or the fact that hSTf contains nineteen disulphide bonds that cannot be formed in *E.coli*.

The fact that these proteins have formed inclusion bodies due to improper folding could be one of the reasons why hSTf expression in *E.coli* was unsuccessful in this study. Since the majority of expression attempts in *E.coli* were carried out using Origami® cells, strains that carry mutations in both the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, the formation of disulfide bonds in the cytoplasm is supposedly 'greatly enhanced'. If disulphide-bond formation is occurring in these cells, there is a chance that correctly folded active hSTf is being expressed within the cytoplasm. However, these active proteins could actually be having a bacteriostatic effect on the cells by chelating all the surrounding iron, inducing a selective pressure that would prevent them from expressing either hSTf or hLf, even as inclusion bodies. Using constructs containing the OmpT leader sequence could also increase this selective pressure, since this fusion leader directs expressed proteins into the periplasm, increasing protein solubility and enhancing protein folding. However, this phenomenon does not explain the lack of expression by the BL21 strains.

Unlike attempts using *E.coli*, hSTf expression was successful using *S.cerevisiae*. Following analysis of the culture supernatant, the transferrin yield for WT rhSTf fermentation was ~1.82mg/mL and for G394R rhSTf fermentation was ~1.7mg/ml. These levels are higher than any previously reported for transferrin, with the closest being by Mizutani *et al*, 2004, who reported levels up to 100mg/L of ovotransferrin using *Pichia pastoris*. However, these levels were achieved in a shake flask and not under the tight control of a fermenter as in these studies.

The purification of recombinant hSTf was achieved using a variety of chromatographic techniques. Initial purification, used two ion-exchange columns, but produced an eluate that still contained contaminants and only retained an average of 54% of the recombinant protein loaded. This loss of hSTf can be attributed to the overloading of the columns and their limit binding capacity. In contrast with the first one, the second purification step, using the immunoaffinity column, was very efficient. By determining a more accurate binding capacity, very little protein was lost and a degree of purity close to 100% was achieved. Such results confirmed that, for research purposes, recombinant hSTf could in the future be purified in a one step process using immunoaffinity chromatography.

Following native PAGE, mass spectrometry and N-terminal sequencing, the pure preparations or recombinant hSTf were shown to comprise of two species. One of these species was fully processed whereas the second one was only partially processed, with a six amino acid residual fragment of the fusion leader on its N-terminal. Paradoxically, the incomplete processing of all expressed proteins could be due to the extreme efficiency of the system at the transcription and translation levels. The presence of these two species could be that the cells express hSTf faster than they are able to process it. In the future, this could possibly be prevented by either reducing the growth rate or, at a genetic level, by experimenting with different fusion leader sequences. For our purpose, the two species were resolved using FPLC MonoQ ion-exchange chromatography.

Even though only ~20% of proteins loaded onto the FPLC MonoQ were recovered a homogeneous sample of each form was achieved. This low level of protein recovery

was due to incomplete resolution of the hSTf forms. In fact, the elution profiles show each fraction to contain varying amounts of both species as the salt gradient increases. This suggests that these proteins are not just separated on the base of their charge differences due to the partially processed leader sequence, but also due to other factors. The heterogeneous glycosylation of the proteins should not have any effect since the glycosylations are O-linked mannoses that do not carry charges. However, these sugars could be interacting in a non-specific manner with the gel matrix due to the age of the column and therefore the build up of material within it.

When the native and recombinant forms of WT and G394R mutant proteins were compared, all structural and functional characteristics explored in this study, suggest that these recombinant proteins, even with partially processed fusion leaders, mimic almost identically their native form. The only difference seen is a 2-fold increase in iron-binding affinity exhibited by the recombinant proteins. Since it is unlikely that the presence of N-linked glycosylations on the native proteins would reduce their affinity for iron, the most likely reason would be experimental error. In fact, the amount of error calculated for the recombinant Kd values is on average significantly higher than the values for native hSTf. Regardless of this, these results imply that the *S.cerevisiae* expression system obtained is an ideal tool for the production of functional hSTf, in both WT and variant forms.

As expected, differences in structural and functional characteristics can be observed when comparing the WT and the G394R forms. Even though there is no significant change in secondary structure, far UV CD spectra suggests that there is a significant change in the tertiary structure of hSTf upon binding to iron with a stronger effect

detected with the G394R mutation. These differences between mutant and WT hSTf are also confirmed by EPR spectroscopy and the blue-shift seen in the scanning UV spectra of the two diferric proteins. By compiling the EPR data from this study with spectra published by Aisen *et al,* in 1978, it can be postulated that these changes are directly related to distortions in the C-lobe's iron-binding site.

Results obtained using light scattering and 6M urea PAGE, show that the presence of R394 has reduced the protein's structural stability. Possible reasons for this could be either the 6-7-fold reduction in the C-lobe Kd, obtained by ITC analysis, or the weakening of residue interactions holding the C-lobe domains together. However, Evans *et al*, 1982 reported that the C-lobe of the iron-free variant is less stable than the C-lobe of the iron-free WT protein, suggesting that the reduction in stability is not directly due to a lowered affinity for iron, but possibly also due to residue interactions elsewhere in the C-lobe. This could also explain why the ITC data show a reduction in the ΔH for variant's C-lobe, since more energy would be produced taking the iron-free C-lobe from its more unstable open form, to its iron-bound, closed form.

The model created in Chapter 3 confirms that R394 is not directly involved in iron binding, nor does it seem to be involved in receptor binding. Evans *et al*, 1994 suggested that residue R394 is interacting with residue D392, one of the iron-binding residues, thus perturbing the iron-binding site. However, these studies suggest that the mutation is causing a much larger effect, not only on the iron-binding site, but on the whole C-lobe, with subtle repercussions on the N-lobe.

Since the ΔS of the variant's C-lobe indicates an increase in domain rigidity and the lowered ΔH indicates an increased number of molecular interactions, rather than R394 interacting with D392, it can be postulated from the hSTf model that the positive side chain of R394 could be interacting with residues away from the iron-binding site and towards the region where domains 1 and 2 hinge in the C-lobe (figure 3.15). These interactions could in turn be restricting the flexibility of the two domains, preventing them from efficiently closing over the iron atom. If the two domains are being obstructed, this could explain the distortion of the tertiary structure seen when the protein binds iron.

Interestingly, it is also around this region of the C-lobe (region 91, Figure 3.21) that hSTf is thought to interact with the helical domain of TfR1 (Cheng *et al,* 2004). This could therefore explain the 10-fold reduced affinity of the variant for the TfR1 (6.25 ± $2.53x10^7$ mol^{-1}) in its diferric form (Young *et al,* 1984a). In fact, the binding affinity for the receptor is similar to that shown for monoferric hSTf ($2.5\text{-}4.6x10^7$ M^{-1}) (Young *et al,* 1984b). The suggestion that the variant's C-lobe is adopting a more open structure, as suggested in section 5.1.3.3 based on the proteins mobility on native PAGE, the interaction with the receptor could be comparable to that of an N-monoferric. However, since the C-lobe of diferric G394R hSTf is not as open as it would be in the apo-form, this could explain why the affinity for the receptor is slightly stronger than that for the N-monoferric.

The proposal that the G394R mutation is perturbing the N-lobe can also be explained by this hypothesis, since the proposed distorted region is in close proximity to where both lobes make contact via a 'cushion' of hydrophobic residues packed between

them. A distortion in the tertiary structure of the C-lobe on binding iron could therefore possibly affect this interaction, causing the subtle alterations in the N-lobe suggested by the thermodynamic data collected during these studies.

To briefly summarise the work presented in this thesis, it can be said that after having established an extremely successful expression system, which is envisaged can be exploited for the expression of numerous other eukaryotic proteins, the comparative analysis and characterisation of both WT and mutant proteins, in combination with the predicted 3D model obtained by computer study, have enabled a step close to be made towards solving the mechanism of iron-uptake and -release by transferrin. Further in depth investigations are necessary to solve the entire mechanism, and the most obvious extension to these studies would be to obtain a high-resolution crystal structure of diferric and apo-hSTf, followed by the elucidation of the structure of the G394R hSTf variant. Unfortunately, the difficulties encountered in the crystallisation of these proteins suggest that it is necessary to explore alternative approaches towards gaining a full understanding of hSTf. One of the main tools now available at the conclusion of our study is the yeast expression system and the access to literally unlimited number of variants obtainable by site-directed mutagenesis. One of the numerous example of such an approach would be to express mutant proteins in an attempt to recreate the same effect as the one observed with G394R variant, but without being in such a close proximity to the iron-binding site. Such results would confirm that residue R394 does not directly interact with the iron-binding site, but causes the C-lobe domains to adapt a less dynamic and open structure. Several mutant variants could also be designed in order to investigate further the inter-dependence of both lobes. For example, the different

thermodynamics obtained for each lobes following ITC analysis, could be further investigated using such variants. Any variation of the data attributed to one lobe following mutation introduced in the other one would confirm and define thermodynamically such interdependence. The availability of specific variants will also be an invaluable tool for the comparative studies of interactions between hSTf and chelators, and their competitive iron binding properties. Variants could also be used to define hSTf/TfR1 interactions, which still remain unclear. Probing the TfR1 peptide library with the variant protein could also help in the interpretation of the variants abnormal properties by seeing whether there is a reduction in the affinity of hSTf to region 91 (Figure 3.21). Defining receptor/protein interactions will be important not only for understanding the mechanism for fundamental and medical purposes, but also as a research tool since hSTf in now widely used as growth factor in cell culture. Indeed, a better understanding of their interactions might provide a better tuning and monitoring of growth.

It is satisfying that, although this work has not provided a complete deciphering of the mechanism of iron-uptake and –release, it has undoubtedly produced the most essential tools for such an aim, and has allowed us to move one step closer to the imminent resolution of the problem.

References

Aisen, P. (1980). The Transferrins. Iron in Biochemistry and Medicine, Vol 2. Jacobs, A. and Worwood, M. Academic Press, London. 87-129.

Aisen, P., Enns, C. and Wessling-Resnick, M. (2001). Chemistry and biology of eukaryotic iron metabolism. The international journal of biochemistry and cell biology. 33: 940-959.

Aisen, P., Leibman, A. and Zweier, J. (1978). Stoichiometric and site characteristics of the binding of iron to human transferrin. Journal of Biological Chemistry. 248: 649-653.

Alcantara, J., Yu, R. H. and Schryvers, A. B. (1993). The region of human transferrin involved in binding to the bacterial transferrin receptors is localised in the C-lobe. Molecular Microbiology. 8: 1135-1143.

Anderson, B. F., Baker, H. M., Dodson, E. J., Norris, G. E., Rumball, S. V., Waters, J. M. and Baker, E. N. (1987). Structure of human lactoferrin at 3.2Å resolution. Proc. Natl. Acad. Sci. USA. 84: 1769-1773.

Andrew, N. C. (2000). Iron Homeostasis: Insights from genetics and animal models. Nature Review Genetics. 1: 208-217.

Archibald, F. (1983). *Lactobacillus plantarum*, an organism no requiring iron. FEMS Microbiology Letters. 19: 29-32.

Arosio, P. and Levi, S. (2002). Ferritin, iron homeostasis, and oxidative damage. Free Radical Biology and Medicine. 33(4): 457-463.

Baldwin, G. S. (1993). Comparison of transferrin sequences from different species. Comparative Biochemistry and Physiology –B: Comparative Biochemistry. 106: 203-218.

Baker, E. N., Anderson, B. F., Baker, H. M., Faber, H. R., Smith, C. A. and Sutherland-Smith, A. J. (1997). Lactoferrin: interactions and biological functions. New Jersey: Humana Press: 177-191.

Baker, E. N. and Lindley, P. F. (1992). New perspectives on the structure functions of transferrin. Journal of Inorganic Biochemistry. 47: 147-160.

Baker, H. M., Anderson, B. F. and Baker. E. N. (2003b). Dealing with iron: common structural principles in proteins that transport iron and heme. PNAS. 100(7): 3579-3583.

Baker, H. M., Baker, C. J., Smith, C. A., Baker, E. N. (2000). Metal substitution in transferrins: specific binding of cerium(IV) revealed by the crystal structure of cerium-substituted human lactoferrin. J. Biol. Inorg. Chem. 5: 692-698.

Baker, H. M., He, Q-Y, Briggs, S. K., Mason, A. B. and Baker, E. N. (2003a). Structural and functional consequences of binding site mutations in transferrin: Crystal structures of the Asp63Glu and Arg124Ala mutants of the N-lobe of human transferrin. Biochemistry. 42: 7084-7089.

Bates, G. W. and Wernicke, J. (1971). The kinetics and mechanisms of iron(III) exchange between chelates and transferrin. The Journal of Biological Chemistry. 246(11): 3679-3685.

Battistuzzi, G., Calzolai, L., Messori, L. and Sola, M. (1995). Metal-induced conformational heterogeneity of transferrins: a spectroscopic study of indium(III) and other metal(III)-substituted transferrins. Biochemical and Biophysical Research Communications. 206(1): 161-170.

Beard, J. and Tobin, B. (2000). Iron status and exercise. Am J Clin Nutr. 72: 594S-7S.

Beguin, Y. (2003). Soluble transferrin receptor for the evaluation of erythropoiesis and iron status. Clinica Chimica Acta. 329: 9-22.

Beutler, E., Gelbart, T., Lee, P., Trevino, R., Fernandez, M. A. and Fairbanks, V. F. (2000). Molecular characterisation of a case of atransferrinemia. Blood. 96(13): 4071-4074.

Binford, J. S. and Foster, J. C. (1974). Calorimetry of the transfer of Fe(3) from nitrilotriacetate to apotransferrin in the presence and in the absence of bicarbonate. J. Biol. Chem. 249: 407-412.

Boulton, I. C., Gorringe, A. R., Gorinsky, B., Retzer, M. D., Schryverys, A. B., Joannou, C. L. and Evans, R. W. (1999). Purified meningococcal transferrin-binding protein B interacts with a secondary, strain-specific, binding site in the N-terminal lobe of human transferrin. Biochem. J. 339: 143-149.

Brunel, F., Ochoa, A., Schaeffer, E., Boissier, F., Guillou, Y., Cereghini, S., Cohen, G. N. and Zakin, M. M. (1988). Interactions of DNA-binding proteins with the 5' region of the human transferrin gene. The Journal of Biological Chemistry. 263(21): 10180-10185.

Carlson, H., Zhang, A-S, Fleming, W. H. and Enns, C. A. (2005). The hereditary haemochromatosis protein, HFE, lowers intracellular iron levels independently of transferrin receptor 1 in TRVb cells. Blood. 105(6): 2564-2570.

Chang, W. W. P., Hobson, C., Bomberger, D. C. and Schneider, L. V. (2005). Rapid separation of protein isoforms by capillary zone electrophoresis with new dynamic coatings. Electrophoresis. 26.

Cheng, Y., Zak, O., Aisen, P., Harrison, S. C. and Walz, T. (2004). Structure of the human transferrin receptor-transferrin complex. Cell. 116: 565-576.

Chinery, S. A. and Hinchliffe, E. (1989). A novel class of vectors for yeast transformation. Curr. Genet. 16(1): 21-5.

Conrad, M. E., Umbreit, J. N. and Moore, E. G. (1999). Iron absorption and transport. Am. J. Med. Sci. 318: 213-229.

Cornelissen, C. N. (2003). Transferrin-iron uptake by gram-negative bacteria. Frontiers in Bioscience. 8: d836-847.

Dayhoff, M.O. (1978) *Observed frequencies of amino acid replacements between closely related proteins. Atlas of Protein Sequence and Structure*, **5**, suppl. 3; National Biomedical Research Foundation, Washington D.C.

De Freitas, J. M. and Meneghini, R. (2001). Iron and its sensitivity balance in the cell. Mutation Research. 475: 153-159.

De Smit, M. H., Hoefkens, P., De Jong, G., Van Duin, J., Van Knippendberg, P. H. and Van Eijk, H. G. (1995). Optimized bacterial production of nonglycosylated human transferrin and its half-molecules. Int. J. Biochem. Cell. Biol. 27(8): 839-850.

Dennis Chasteen, N. and Harrison, P. M. (1999). Mineralisation in ferritin: An efficient means of iron storage. Journal of Structural Biology. 126: 182-194.

Donovan, J. W. and Ross, K. D. (1976). Calorimetric studies of the binding of iron and aluminium to human serum transferrin. Fed. Proc. Fed. Am. Soc. Exp. Biol. 35: 1608.

Du, X., Zhang, T., Yuan, L., Zhao, Y., Li, R., Wang, K., Yan, S. C., Zhang, L., Sun, H. and Qian, Z. (2002). Complexation of ytterbium to human transferrin and its uptake by K562 cells. Eur. J. Biochem. 269: 6082-6090.

Erlitzki, R., Long, J. C. and Theil, E. C. (2002). Multiple, conserved iron responsive elements in the 3' translated region of transferrin receptor mRNA enhance binding of iron regulatory protein 2. J. Biol. Chem. 277: 42579-42587.

Evans, R. W., Crawley, J. B., Garratt, R. C., Gunter Grossmann, J., Neu, M., Aitken, A., Patel, K. J., Meilak, A., Wong, C., Singh, J., Bomford, A, and Samar Hasnain, S. (1994). Characterisation and structural analysis of a functional human serum transferrin variant and implications for receptor recognition. Biochemistry. 33(41): 12512-12520.

Evans, R. W. and Madden, A. D. (1984). A low molecular weight iron-binding fragment from duck ovotransferrin. Biochemical Society Transactions. 12: 661-662.

Evans, R. W., Aitken, A. and Patel, K. J. (1988). Evidence for a single glycan moiety in rabbit serum transferrin and location of the glycan within the polypeptide chain. FEBS Lett 238(1):39-42

Evans, R. W. and Williams, J. (1978). Studies of the binding of different iron donors to human serum transferrin and isolation of iron-binding fragments from the N- and C-terminal regions of the protein. Biochem. J. 173: 543-552.

Evans, R. W., Williams, J. and Moreton, K. (1982). A variant of human transferrin with abnormal properties. Biochem J. 201: 19-26.

Garnier, J., Gibrat, J-F. and Robson, B. (1996). GOR secondary structure prediction method version IV. Methods in Enzymology. 266: 540-553.

Garnier, J., Osguthorpe, D. and Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120 (1): 97-120.

Ghareeb, B. A. A., Thepot, D., Delville-Giraud, C. and Houdebine, L-M. (1998). Cloning and functional expression of the rabbit transferrin gene promoter. Gene. 211: 301-310.

Giannetti, A. M. and Björkman, P. J. (2004). HFE and transferrin directly compete for transferrin receptor in solution and at the cell surface. The Journal of Biological Chemistry. 279(24): 25866-25875.

Giannetti, A. M., Snow, P. M., Zak, O. and Björkman, P. J. (2003). Mechanism for multiple ligand recognition by the human transferrin receptor. Plos Biology. 1(3): 341-350.

Gorringe, A. R. and Oakhill, J. (2002). Neisseria meningitidis: multiple mechanisms to acquire iron. Ferrieros, C., Criado, M. T. and Vazquez, J. Horizon Scientific Press.

Gorinsky, B., Horsburgh, C., Lindley, P. F., Moss, D. S., Parkar, M. and Watson, J. L. (1979). Evidence for the bilobal nature if diferric rabbit plasma transferrin. Nature. 281: 157-158.

Guha Thakurta, P., Choudhury, D., Dasgupta, R. and Dattagupta, J. K. (2003). Structure of diferric hen serum transferrin at 2.8Å resolution. Acta Cryst. D59: 1773-1781.

Guha Thakurta, P., Choudhury, D., Dasgupta, R. and Dattagupta, J. K. (2004). Tertiary structural changes associated with iron binding and release in hen serum transferrin: a crystallographic and spectroscopic study. Biochemical and biophysical research communications. 316: 1124-1131.

Gumerov, D. R., Mason, A. B. and Kaltashov, I. A. (2003). Interlobe communication in human serum transferrin: metal binding and conformational dynamics investigated by electrospray ionisation mass spectrometry. Biochemistry. 42: 5421-5428.

Gunshin, H., Mackenzie, B. and Berger, U. V. (1997). Cloning and characterization of a mammalian proton coupled metal-ion transporter. Nature. 388: 482-488.

Hall, D. R., Hadden, J. M., Leonard, G. A., Bailey, S., Neu, M., Winn, M. and Lindley, P. F. (2001). The crystal and molecular structures of diferric porcine and rabbit serum transferrins at resolutions of 2.15 and 2.6Å, respectively. Acta Cryst. D58: 70-80.

Haridas, M., Anderson, B. F. and Baker, E. N. (1995). Structure of human diferric lactoferrin refined at 2.2Å resolution. Acta Cryst. D51: 629-646.

Harris, M., Taylor, G. and Taylor, J. (2005). Maths and Stats: For the life and medical sciences. CatchUp. Scion Publishing Limited.

Harris, W. R., Yang, B., Abdollahi, S. and Hamada, Y. (1999). Steric restrictions on the binding of large metal ions to serum transferrin. J. Inorg. Biochem. 76: 231

Hausler, A., Ballou, L., Ballou, C. E. and Robbins, P. W. (1992). Yeast glycoproteins biosynthesis: MNT1 encodes an α-1,2-mannosyltransferase involved in O-glycosylation. Proc. Natl. Acad. Sci. USA. 89: 6846-6850.

He, Q-Y., Mason, A. B., Lyons, B. A., Tam, B. M., Nguyen, V., MacGillivray, R. T. A. and Woodworth, R. C. (2001). Spectral and metal-binding properties of three

single-point tryptophan mutants of the human transferrin N-lobe. Biochem. J. 354: 423-429.

He, Q-Y., Mason, A. B., Pakdaman, R., Chasteen, N. D., Dixon, B. K., Tam, B. M., Nguyen, V., MacGillivray, R. T. A. and Woodworth, R. C. (2000). Mutations at the histidine 249 ligand profoundly alter the spectral and iron-binding properties of human serum transferrin N-lobe. Biochemistry 39: 1205-1210.

He, Q-Y., Mason, A. B., Tam, B. M., MacGillivray, R. T. A. and Woodworth, R. C. (1999). [13C]Methionine NMR and metal-binding studies of recombinant human transferrin N-lobe and five methionine mutants: conformational changes and increased sensitivity to chloride. Biochem. J. 344: 881-887.

He, Q-Y., Mason, A. B., Woodworth, R. C., Tam, B. M., MacGillivray, R. T. A., Grady, J. K. and Dennis Chasteen, N. (1998). Mutations at nonliganding residues Tyr-85 and Glu-83 in the N-lobe of human serum transferrin. The Journal of Biological Chemistry. 273(27): 17018-17024.

Hentze, M. W. and Kühn, L. C. (1996). Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. Proc. Natl. Acad. Sci. USA. 93: 8175-8182.

Hentze, M. W., Muckenthaler, M. U. and Andrews, N. C. (2004). Balancing acts: Molecular control of mammalian iron metabolism. Cell. 117: 285-297.

Hopp, T. P. and Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA. 78(6): 3824-3828.

Idzerda, R. L., Behrindger, R. R., Theisen, M., Guggenvik, J. I., McKinght, G. S. and Brinster, R. L. (1989). Expression from the transferrin gene promoter in transgenic mice. Molecular and Cellular Biology. 9(11): 5154-5162.

Idzerda, R. L., Huebers, H., Finch, C. A. and McKnight, G. S. (1986). Rat transferrin gene expression: Tissue-specific regulation by iron deficiency. Proc. Natl. Acad. Sci. USA. 83: 3723-3727.

Ikemoto, H. and Ventura, M. N. (1979). Differential scanning calorimetry of the thermal denaturation of human serotransferrin. An. Acad. Bras. Cienc. 51(1): 165-71.

Jameson, G. B., Anderson, B. F., Norris, G. E., Thomas, D. H. and Baker, E. N. (1998). Structure of human apolactoferrin at 2.0Å resolution. Refinement and analysis of ligand-induced conformational change. Acta. Cryst. D54: 1319-1335.

Jamieson, G. A., Jett, M. and DeBernardo, S. L. (1971). The carbohydrate sequence of the glycopeptide chains of human transferrin. The Journal of Biological Chemistry. 246(11): 3686-3693.

Janin, J. (1979). Surface and inside volumes in globular proteins. Nature. 277(5696): 491-492.

Jelesarov, I. and Bosshard, H. R. (1999). Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. Journal of Molecular Recognition. 12: 3-18.

Karthikeyan, S., Yadav, S., Paramasivam, M. and Singh, T. P. (2000). Structure of buffalo lactoferrin at 3.3Å resolution at 277K. Acta Cryst. D56: 684-689.

Kawabata, H., Yang, R., Hirama, T., Vuong, P. T., Kawano, S., Gombart, A. F. and Koeffler, H. P. (1999). Molecular cloning of transferrin receptor 2. J. Biol. Chem. 274: 20826-20832.

Kelly, S. M., and Price, N. C. (2000). The use of circular Dichroism in the investigation of protein structure and function. Curr Protein Pept Sci. 1(4): 349-84.

Kerry-Williams, S. M., Gilbert, S. C., Evans, L. R. and Balance, D. J. (1998). Disruption of the *Saccharomyces cerevisae YAP3* gene reduces the proteolytic degradation of secreted recombinant human albumin. Yeast, 14: 6-169.

Khan, J. A., Kumar, P., Paramasivam, M., Yadav, R. S., Sahani, M. S., Sharma, S., Srinivasan, A. and Singh, T. P. (2001). Camel lactoferrin, a Transferrin-cum-lactoferrin: Crystal structure of camel lactoferrin at 2.6Å resolution and structural basis of its dual role. J. Mol. Biol. 309: 751-761.

Kihara, D. (2005). The effect of long-range interactions on the secondary structure formation of proteins. Protein Science. 14: 1955-1963.

Kumar, P., Khan, J. A., Yadav, S. and Singh, T. P. (2002). Crystal structure of equine apolactoferrin at 303K providing further evidence of closed conformations of N and C lobes. Acta Cryst. D58: 225-232.

Kyte, J. (1995). Structure in protein chemistry. Garland Publishing, Inc. New York and London.

Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157(1): 105-32.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 227(259): 680-5

Lane, R. S. (1975). Differences between human Fe1-transferrin molecules. British Journal of Haematology. 29: 511-520.

Lawrence, C. M., Ray, S., Babyonsyhev, M., Galluser, R., Borhani, D. W. and Harrison, S. C. (1999). Crystal structure of the ectodomain of human transferrin receptor. Science. 286: 779-782.

Lee, P. L., Halloran, C., Trevino, R., Felitti, V. and Beutler, E. (2001). Human transferrin G277S mutation: a risk factor for iron deficiency anaemia. British Journal of Haematology. 115: 329-333.

Lee, P. L., Ho, N. J., Olson, R. and Beutler, E. (1999). The effects of transferrin polymorphisms on iron metabolism. Blood Cells, Molecules and Diseases. 25(24): 374-379.

Lehrer, S. S. (1969). Fluorescence and absorption studies of the binding of copper and iron to Transferrin. The Journal of Biological Chemistry. 244(13): 3613-3617.

Li, H., Sadler, P. J., Sun, H. (1996) Unexpectedly strong binding of a large metal ion (Bi3+) to human serum transferrin. J. Biol. Chem. 271: 9483-9489.

Lieu, P. T., Heiskala, M., Peterson, P. A. and Yang, Y. (2001). The roles of iron in health and disease. Molecular Aspects of Medicine. 22: 1-87.

Lin, L-N., Mason, A. B., Woodworth, R. C. and Brandts, J. F. (1993). Calorimetric studies of the binding of ferric ions to human serum transferrin. Biochemistry. 32: 9398-9406.

MacGillivray, R. T. A., Mendez, E., Shewale, J. G., Sinha, S. K., Lineback-Zins, J. and Brew, K. (1983). The primary structure of human serum transferrin. The journal of biological chemistry. 258, 6: 3543-3553.

MacGillivray, R. T. A., Mendez, E., Sinha, S. K., Sutton, M. R., Lineback-Zins, J. and Brew, K. (1982). The complete amino acid sequence of humans serum transferrin. Proc. Natl. Acad. Sci. USA. 79: 2504-2508.

MacGuffin. L. J., Bryson. K. and Jones, D. T. (2000). The PSIPRED protein structure prediction server. Bioinformatics. 16: 404-405.

Makey, D. G. and Seal, U. S. (1976). The detection of four molecular forms of human transferrin during the iron binding process. Biochem. Biophys. Acta. 453(1): 250-256.

Malik, K. M. (1997). Chemometric and quantum mechanical methods to analyse CD data. Thesis. King's College London, University of London.

Mason, A. B., Miller, M. K., Funk, W. D., Banfield, D. K., Savage, K. J., Oliver, R. W., Green, B. N., MacGillivray, R. T. and Woodworth, R. C. (1993). Expression of glycosylated and nonglycosylated human transferrin in mammalian cells. Characterisation of the recombinant proteins with comparison to three commercially available transferrins. Biochemistry. 32(20): 5472-9.

Mason, A. B., Woodworth, R. C., Oliver, R. W. A., Green, B. N., Lin, L-N., Brandts, J. F., Savage, K. J., Tam, B. M. and MacGillivray, R. T. A. (1996). Association of the two lobes of ovotransferrin is a prerequisite for receptor recognition. Biochem. J. 319: 361-368.

McGraw, T. E., Greenfield, L. and Maxfield, F. R. (1987). Functional expression of the human transferrin receptor cDNA in Chinese hamster ovary cells deficient in endogenous transferrin receptor. J Cell Biol. 105(1), 207-14.

Meyron-Holtz, E. G., Ghosh, M. C., Iwai, K., LaVaute, T., Brazzolotto X., Berger, U. V., Land, W., Ollivierre-Wilson, H., Grinberg, A., Love, P. and Rouault, T. A. (2004). Genetic ablations of iron regulatory proteins 1 and 2 reveals why iron regulatory protein 2 dominates iron homeostasis. The EMBO Journal. 23: 386-395.

Mijajima, H. (2003). Aceruloplasminemia, an iron metabolic disorder. Neuropathology. 23: 345-350.

Mizutani, K., Mikami, B. and Hirose, M. (2001). Domain closure mechanism in transferrin: New viewpoints about the hinge structure and motion as deduced from high resolution crystal structures of ovotransferrin N-lobe. J. Mol. Biol. 309: 937-947.

Mizutani, K., Okamoto, I., Fujita, K., Yamamoto, K. and Hirose, M. (1994). Structural and functional characterisation of ovotransferrin produced by *Pichia pastoris*. Biosci. Biotechnol, Biochem. 68(2): 376-383.

Mok, H., Jelinek, J., Pai, S., Cattanach, B. M., Prchal, J. T., Youssoufian, H. and Schumacher, A. (2004). Disruption of ferroportin 1 regulation causes dynamic alterations in iron homeostasis and erythropoiesis in polycythaemia mice. Development and Disease. 131: 1859-1868.

Moore, S. A., Anderson, B. F., Groom, C. R., Haridas, M. and Baker, E. N. (1997). Three-dimensional structure of diferric bovine lactoferrin at 2.8Å resolution. J. Mol. Biol. 274: 222-236.

Moriarty, L. C. (2002). The physiological relevance of lactoferrin and other microbial peptides of lactoferrin. Thesis. School of Crystallography, Birkbeck College, London.

Morgan, E. H. (1996). Cellular iron processing. J. Gastroenterol. Hepatol. 11:1027-1030.

Nurizzo, D., Baker, H. M., He, Q-Y., MacGillivray, R. T. A., Mason, A. B., Woodworth, R. C. and Baker, E. N. (2001). Crystal structures and iron release properties of mutants (K206A and K296A) that abolish the dilysine interactions in the N-lobe of human transferrin. Biochemistry. 40: 1616-1623.

Osborne, T. B. and Campbell, G. F. (1900). The protein constituents of egg white. Journal of American Chemical Society. 22: 422-449.

Pace, C. N. (1990). Measuring and increasing protein stability. Trends Biotechnol. 8(4):93-8.

Pantopoulos, K. (2004). Iron metabolism and the IRE/IRP regulatory system. Ann. N. Y. Acad. Sci. 1012: 1-13.

Park, I., Schaeffer, L., Sidoli, E., Baralle, F. E., Cohen, G. N. and Zakin, M. M. (1985). Organisation of the human serum transferrin gene: direct evidence that it originated by gene duplication. Proc. Natl. Acad. Sci. USA. 82(10): 3149-3153.

Pierce, M. M., Raman, C. S. and Nall, B. T. (1999). Isothermal Titration Calorimetry of protein-protein interactions. Methods. 19: 213-221.

Pinkowitzh, R. A. and Aisen, P. (1972). Zero-field splitting of iron complexes of transferrins. The Journal of Biological Chemistry. 247(23): 7830-7834.

Popovic, Z. and Templeton, D. M. (2004). Iron accumulation and iron-regulatory protein activity in human hepatoma (HepG2) cells. Mol. Cell. Biochem. 265(1-2): 37-45.

Pulina, M. O., Zakharova, E. T., Sokolov, A. V., Shavlovski, M. M., Bass, M. G., Solovyov, K. V., Kokryakov, V. N. and Vasilyev, V. B. (2002). Studies of the ceruloplasmin-lactoferrin complex. Biochemistry and Cell Biology. 80(1):25-39.

Qian, Z. M., Li, H., Sun, H. and Ho, K. (2002). Targeted drug delivery via the transferrin receptor-mediated endocytosis pathway. Pharmacological Reviews. 54(4): 561-587.

Ratledge, C. and Dover, L. G. (2000). Iron metabolism in pathogenic bacteria. Annu. Rev. Microbiol. 54: 881-941.

Retzer, M. D., Yu, R. and Schryvers, A. B. (1999). Identification of sequences in human transferrin that bind to the bacterial receptor protein, transferrin-binding protein B. Molecular Microbiology. 32(1): 111-121.

Rinaldo, D. and Field, M. J. (2003). A computational study of the open and closed forms of the N-lobe human serum transferrin apoprotein. Biophysical Journal. 85: 3485-3501.

Rosalki, S. B. (2004). Carbohydrate-deficient transferrin: a marker of alcohol abuse. Int J Clin Pract. 58(4): 391-393.

Roy, C. N. and Andrews, N. C. (2001). Recent advances in disorders of iron metabolism: mutations, mechanisms and modifiers. Human Molecular Genetics. 10(20): 2181-2186.

Roy, C. N., Penny, D. M., Feder, J. N. and Enns, C. A. (1999). The hereditary hemochromatosis protein, HFE, specifically regulates transferrin-mediated iron uptake in HeLa cells. The Journal of Biological Chemistry. 274(13): 9022-9028.

Sali, A., Fiser, A., Sanchez, R., Marti-Renom, M. A., Jerkovic, B., Badretdinov, A., Melo, F., Overington, J. P. and Feyfant, E. (2001). Modeller. A Program for Protein Structure Modelling. Release 6. The Rockefeller University, New York.

Sali, A., Potterton, L., Yuan, F., van Vlijmen, H. and Karplus, M. (1995). Evaluation of comparative protein modelling by MODELLER. Proteins. 23(3): 318-26.

Sarra, R., Garratt, R., Gorinsky, B., Jhoti, H. and Lindley, P. (1990). High-resolution X-ray studies on rabbit serum transeferrin: preliminary structure analysis of the N-terminal half-molecule at 2.3Å resolution. Acta. Cryst. B46: 763-771.

Schade, A. L. and Caroline, L. (1944). Raw hen egg white and the role of iron in growth inhibition of *Shigella dysenteriae, Staphylococcus aureus, Escherichia coli* and *Saccharomyces cerevisiae*. Science. 100: 14-15.

Schade, A. L. and Caroline, L. (1946). An iron binding component in human blood plasma. Science. 104: 340-341.

Schaeffer, E., Guillou, F., Part, D. and Zakin, M. M. (1993). A different combination of transcription factors modulates the expression of the human transferrin promoter in liver and Sertoli cells. The Journal of Biological Chemistry. 268(31): 23399-23408.

Sharma, A. K. and Singh, T. P. (1999). Lactoferrin-metal interactions: first crystal structure of a complex of lactoferrin with a lanthanide ion (Sm3+) at 3.4 A resolution. Acta Cryst. D. 55: 1799-1804.

Shen, Z. M., Yang, J. T., Feng, Y-M. and Wu, C-S. C. (1992). Conformational stability of porcine serum transferrin. Protein Science. 1: 1477-1484.

Sheth, S. and Brittenham, G. M. (2000). Genetic disorders affecting proteins of iron metabolism: clinical implications. Annu. Rev. Med. 51: 443-464.

Sleep, D., Belfield, G. P., Balance, D. J., Steven, J., Jones, S., Evans, L. R., Moir, P. M. and Goodey, A. R. (1991). *Saccharomyces cerevisiae* strains that over express heterologous proteins. Bio/Technology, 9: 183-187.

Sleep, D., Belfield, G. P. and Goodey, A. R. (1990). The secretion of human serum albumin from the yeast *Saccharomyces cerevisiae* using five different leader sequences. Bio/Technology, 8: 42-46.

Smith, C. A., Ainscough, E. W., Baker, H. M., Brodie, A. M., Baker, E. N. (1994). Specific binding of cerium by human lactoferrin stimulates the oxidation of Ce^{3+} to $Ce4^{+}$. J. Am. Chem. Soc. 116: 7889-7890.

Smith, C. A., Anderson, B. F., Baker, H. M. and Baker, E. N. (1992). Metal substitution in transferrins: the crystal structure of human copper-lactoferrin at 2.1-A resolution. Biochemistry. 31: 4527-4533.

Steinlein, L. M. and Ikeda, R. A. (1993). Production of N-terminal and C-terminal human serum transferrin in *Escherichia coli*. Enzyme Microb Technol. 15: 193-199.

Steinlein, L. M., Ligman, C. M., Kessler, S. and Ikeda, R. A. (1998). Iron release in reduced by mutations of lysines 206 and 296 in recombinant N-terminal half-transferrin. Biochemistry. 37(39): 13696-703.

Sutton, R., Rockett, B. and Swindells, P. (2000). Chemistry for the life sciences. Life Lines. Taylor & Franscis.

Tapiero, H., Gate, L. and Tew, K. D. (2001). Iron deficiencies and requirements. Biomed Pharmacother. 55: 324-32.

Thakurta, P. G., Choudhury, D., Dasgupta, R. and Dattagupta, J. K. (2004). Tertiary structural changes associated with iron binding and release in hen serum transferrin: a crystallographic and spectroscopic study. Biochemical and Biophysical Research Communications. 316: 1124-1131.

Torti, F. M. and Torti, S. V. (2002). Regulation of ferritin genes and protein. Blood. 99(10): 3505-1516.

Trinder, D. and Baker, E. (2003). Transferrin receptor 2: a new molecule in iron metabolism. The International Journal of Biochemistry and Cell Biology. 35: 292-296.

Trowbrigdge, I. S. and Shackelford, D. A. (1986). Structure and function of transferrin receptors and their relationship to cell growth. Biochem. Soc. Symp. 51: 117-29.

Ugo Testa. (2002). Proteins of iron metabolism. CRC Press.

Van den Steen, P., Rudd, P. M., Dwek, R. A., Opdenakker, G. (1998). Concepts and principles of O-linked glycosylation. Critical Reviews in Biochemistry and Molecular Biology. 33 (3): 151-208.

Vyhlidal, C. and Safe, S. (2002). Estrogen regulation of transferrin gene expression in MCF-7 human breast cancer cells. Journal of Molecular Endocrinology. 29: 305-317.

Wallace, B. A. and Janes, R. W. (2001). Synchrotron radiation circular dichroism spectroscopy of proteins: secondary structure, fold recognition and structural genomics. Current Opinion in Chemical Biology. 5: 567-571.

Wang, J. and Pantopoulos, K. (2002). Conditional derepression of ferritin synthesis in cell expressing a constitutive IRP1 mutant. Molecular and Cell Biology. 22(13): 4638-4651.

Wardrop, S. L. and Richardson, D. R. (1999). The effect of intracellular iron concentration and nitrogen monoxide on Nramp2 expression and non-transferrin-bound iron uptake. Eur. J. Biochem. 263: 41-49.

Welch, S. (1992). Transferrin: The iron carrier. CRC Press: Boca Raton.

Westhead, D. R., Parish, J. H. and Twyman, R. M. (2002). Instant Notes: Bioinformatics. Bios Scientific Publishers Limited.

Williams, J. (1982). The evolution of transferrin. Trends in Biochemical Sciences. 7: 394-397.

Williams, J., Evans, R. W. and Moreton, K. (1978). The iron-binding properties of hen ovotransferrin. Biochem. J. 173(2): 533-539.

Wilson, K. and Walker, J. (2000). Principles and techniques of practical biochemistry. Fifth edition. Cambridge University Press.

Wong, H. and Schryvers, A. B. (2003). Bacterial lactoferrin-binding protein A binds to both domains of the human lactoferrin C-lobe. Microbiology. 149: 1729-1737.

Woodworth, R. C., Mason, A. B., Funk, W. D. and MacGillivray, R. T. A. (1991). Expression and initial characterisation of five site-directed mutants of the N-terminal half-molecule of human transferrin. Biochemistry. 1991. 30(45): 10824-9.

Xiong, S., Wang, Y. F., Ren, X. R., Li, B., Zhang, M. Y., Luo, Y., Zhang, L., Xie, Q. L. and Su, K. Y. (2005). Solubility of disulfide-bonded proteins in the cytoplasm of Escherichia coli and its "oxidizing" mutant. World J Gastroenterol. 11(7): 1077-1082.

Yang, A. H. W., MacGillivray, R. T. A., Chen, J., Luo, Y., Wang, Y., Brayer, G. D., Mason, A. B., Woodworth, R. C. and Murphy, M. E. P. (2000). Crystal structures of two mutants (K206Q, H207E) of the N-lobe of human transferrin with increased iron affinity for iron. Protein Science. 9: 49-52.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene. 33(1): 103-19.

Young, S. P., Bomford, A., Madden, A. D., Garratt, R. C., Williams, R. and Evans R. W. (1984a). Abnormal in vitro function of a variant human transferrin. Br. J. Haemotol. 56(4): 581-7.

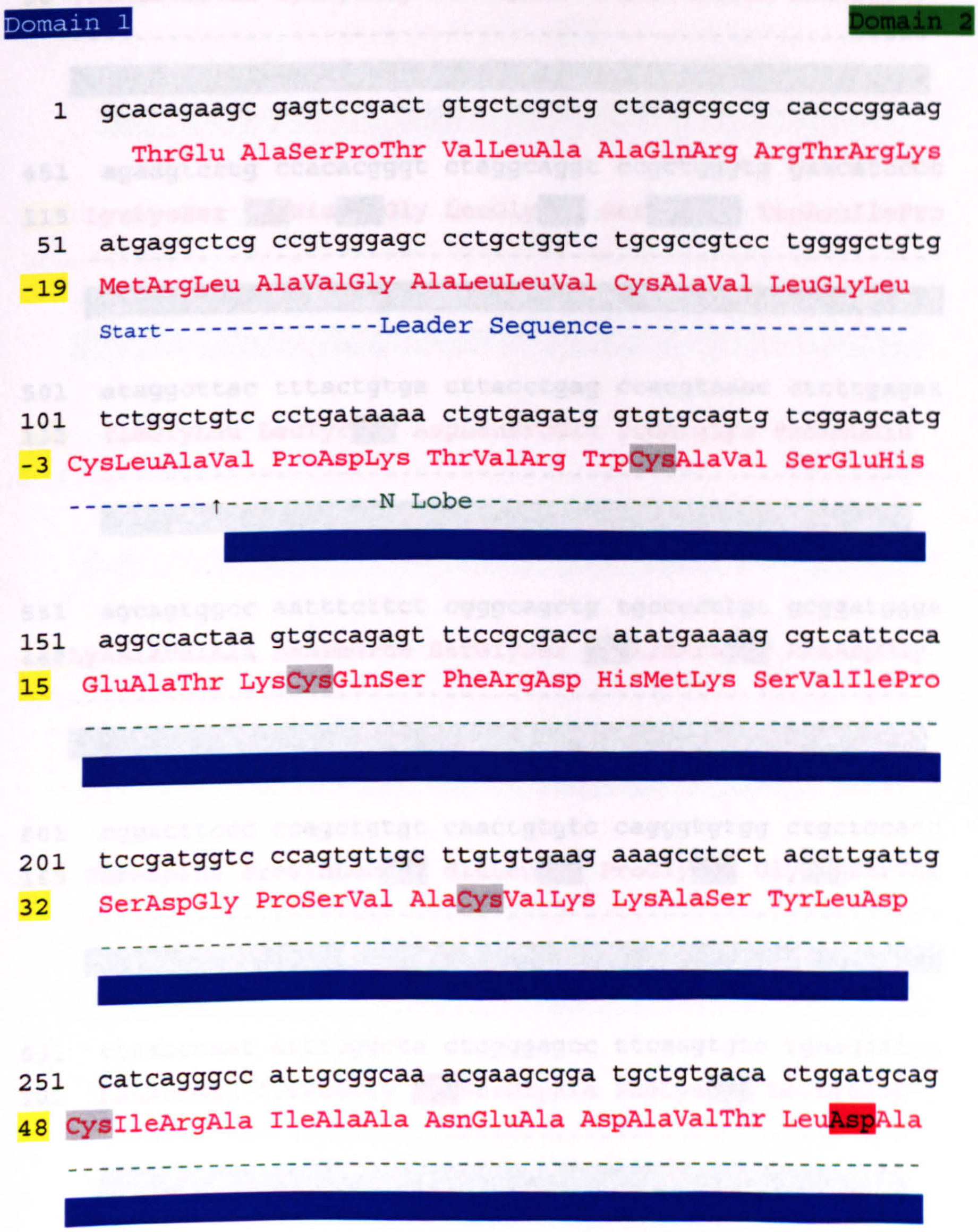
Young, S. P., Bomford, A. and Williams, R. (1984b). The effect of the iron saturation of transferrin on its binding and uptake by rabbit reticulocytes. Biochem. J. 219: 505-510.

Zhang, M. Gumerov, D. R., Kaltashov, I. A. and Mason, A. B. (2004). Indirect detection of protein-metal binding: interaction of serum transferrin with In3+ and Bi3+. J. Am. Soc. Mass. Spectrom. 15: 1658-1664.

Appendix I: Computational Analysis

I.1. Human Serum Transferrin Sequence

The sequence of hSTf with the leader, the two lobes, the bridge and the two domains highlighted. Also highlighted, are the 38 cysteines thought to be involved in disulphide-bridge formation (grey), the eight residues theorised to be involved in Fe^{3+} binding (red) and the eight residues thought to be involved in the carbonate binding (turquoise).



```
301  gtttggtgta tgatgcttac ctggctccca ataacctgaa gcctgtggtg
65  GlyLeuVal TyrAspAlaTyr LeuAlaPro AsnAsnLeu LysProValVal
    ------------------------------------------------------
```

```
351  gcagagttct atgggtcaaa agaggatcca cagactttct attatgctgt
82   AlaGluPhe TyrGlySer LysGluAspPro GlnThrPhe TyrTyrAla
     ------------------------------------------------------
```

```
401  tgctgtggtg aagaaggata gtggcttcca gatgaaccag cttcgaggca
98  ValAlaValVal LysLysAsp SerGlyPhe GlnMetAsnGln LeuArgGly
    ------------------------------------------------------
```

```
451  agaagtcctg ccacacgggt ctaggcaggt ccgctgggtg gaacatcccc
115 LysLysSer CysHisThrGly LeuGlyArg SerAlaGly TrpAsnIlePro
    ------------------------------------------------------
```

```
501  ataggcttac tttactgtga cttacctgag ccacgtaaac ctcttgagaa
132  IleGlyLeu LeuTyrCys AspLeuProGlu ProArgLys ProLeuGlu
     ------------------------------------------------------
```

```
551  agcagtggcc aatttcttct cgggcagctg tgcccttgt gcggatggga
148LysAlaValAla AsnPhePhe SerGlySer CysAlaProCys AlaAspGly
    ------------------------------------------------------
```

```
601  cggacttccc ccagctgtgt caactgtgtc cagggtgtgg ctgctccacc
165 ThrAspPhe ProGlnLeuCys GlnLeuCys ProGlyCys GlyCysSerThr
    ------------------------------------------------------
```

```
651  cttaaccaat acttcggcta ctcgggagcc ttcaagtgtc tgaaggatgg
182  LeuAsnGln TyrPheGly TyrSerGlyAla PheLysCys LeuLysAsp
     ------------------------------------------------------
```

```
701   tgctggggat gtggcctttg tcaagcactc gactatattt gagaacttgg
198GlyAlaGlyAsp ValAlaPhe ValLysHis SerThrIlePhe GluAsnLeu
```

```
751   caaacaaggc tgacagggac cagtatgagc tgctttgcct ggacaacacc
215 AlaAsnLys AlaAspArgAsp GlnTyrGlu LeuLeuCys LeuAspAsnThr
```

```
801   cggaagccgg tagatgaata caaggactgc cacttggccc aggtcccttc
232   ArgLysPro ValAspGlu TyrLysAspCys HisLeuAla GlnValPro
```

```
851   tcataccgtc gtggcccgaa gtatgggcgg caaggaggac ttgatctggg
248SerHisThrVal ValAlaArg SerMetGly GlyLysGluAsp LeuIleTrp
```

```
901   agcttctcaa ccaggcccag gaacattttg gcaaagacaa atcaaaagaa
265 GluLeuLeu AsnGlnAlaGln GluHisPhe GlyLysAsp LysSerLysGlu
```

```
951   ttccaactat tcagctctcc tcatgggaag gacctgctgt ttaaggactc
282   PheGlnLeu PheSerSer ProHisGlyLys AspLeuLeu PheLysAsp
```

```
1001  tgcccacggg tttttaaaag tccccccccag gatggatgcc aagatgtacc
298 SerAlaHisGly PheLeuLys ValProPro ArgMetAspAla LysMetTyr
```

```
1051  tgggctatga gtatgtcact gccatccgga atctacggga aggcacatgc
315   LeuGlyTyr GluTyrValThr AlaIleArg AsnLeuArg GluGlyThrCys
                                                          ↑==
```

```
1101  ccagaagccc caacagatga atgcaagcct gtgaagtggt gtgcgctgag
332     ProGluAla ProThrAsp GluCysLysPro ValLysTrp CysAlaLeu
        =Bridge Region============↑----------C Lobe---------
```

```
1151  ccaccacgag aggctcaagt gtgatgagtg gagtgttaac agtgtaggga
348  SerHisHisGlu ArgLeuLys CysAspGlu TrpSerValAsn SerValGly
     ----------------------------------------------------
```

```
1201  aaatagagtg tgtatcagca gagaccaccg aagactgcat cgccaagatc
365   LysIleGlu CysValSerAla GluThrThr GluAspCys IleAlaLysIle
      ----------------------------------------------------
```

```
1251  atgaatggag aagctgatgc catgagcttg gatggagggt ttgtctacat
382     MetAsnGly GluAlaAsp AlaMetSerLeu AspGlyGly PheValTyr
        ----------------------------------------------------
```

```
1301  agcgggcaag tgtggtctgg tgcctgtctt ggcagaaaac tacaataaga
398  IleAlaGlyLys CysGlyLeu ValProVal LeuAlaGluAsn TyrAsnLys
     ----------------------------------------------------
```

```
1351  gcgataattg tgaggataca ccagaggcag ggtattttgc tgtagcagtg
415   SerAspAsn CysGluAspThr ProGluAla GlyTyrPhe AlaValAlaVal
      ----------------------------------------------------
```

```
1401  gtgaagaaat cagcttctga cctcacctgg gacaatctga aaggcaagaa
432     ValLysLys SerAlaSer AspLeuThrTrp AspAsnLeu LysGlyLys
        ----------------------------------------------------
```

```
1451  gtcctgccat acggcagttg gcagaaccgc tggctggaac atccccatgg
448  LysSerCysHis ThrAlaVal GlyArgThr AlaGlyTrpAsn IleProMet
     ----------------------------------------------------
```

```
1501  gcctgctcta caataagatc aaccactgca gatttgatga atttttcagt
465   GlyLeuLeu TyrAsnLysIle AsnHisCys ArgPheAsp GluPhePheSer
      ----------------------------------------------------
```

```
1551  gaaggttgtg cccctgggtc taagaaagac tccagtctct gtaagctgtg
482     GluGlyCys AlaProGly SerLysLysAsp SerSerLeu CysLysLeu
        ------------------------------------------------
```

```
1601  tatgggctca ggcctaaacc tgtgtgaacc caacaacaaa gagggatact
498   CysMetGlySer GlyLeuAsn LeuCysGlu ProAsnAsnLys GluGlyTyr
      ----------------------------------------------------
```

```
1651  acggctacac aggcgctttc aggtgtctgg ttgagaaggg agatgtggcc
515   TyrGlyTyr ThrGlyAlaPhe ArgCysLeu ValGluLys GlyAspValAla
      ----------------------------------------------------
```

```
1701  tttgtgaaac accagactgt cccacagaac actgggggaa aaaaccctga
532     PheValLys HisGlnThr ValProGlnAsn ThrGlyGly LysAsnPro
        ------------------------------------------------
```

```
1751  tccatgggct aagaatctga atgaaaaaga ctatgagttg ctgtgccttg
548   AspProTrpAla LysAsnLeu AsnGluLys AspTyrGluLeu LeuCysLeu
      ----------------------------------------------------
```

```
1801  atggtaccag gaaacctgtg gaggagtatg cgaactgcca cctggccaga
565   AspGlyThr ArgLysProVal GluGluTyr AlaAsnCys HisLeuAlaArg
      ----------------------------------------------------
```

```
1851  gccccgaatc acgctgtggt cacacggaaa gataaggaag cttgcgtcca
582     AlaProAsn HisAlaVal ValThrArgLys AspLysGlu AlaCysVal
        ------------------------------------------------
```

```
1901  caagatatta cgtcaacagc agcacctatt tggaagcaac gtaactgact
598 HisLysIleLeu ArgGlnGln GlnHisLeu PheGlySerAsn ValThrAsp
    ----------------------------------------------------
```

```
1951  gctcgggcaa cttttgtttg ttccggtcgg aaaccaagga ccttctgttc
615   CysSerGly AsnPheCysLeu PheArgSer GluThrLys AspLeuLeuPhe
      ----------------------------------------------------
```

```
2001  agagatgaca cagtatgttt ggccaaactt catgacagaa acacatatga
632     ArgAspAsp ThrValCys LeuAlaLysLeu HisAspArg AsnThrTyr
        ----------------------------------------------------
```

```
2051  aaaatactta ggagaagaat atgtcaaggc tgttggtaac ctgagaaaat
648 GluLysTyrLeu GlyGluGlu TyrValLys AlaValGlyAsn LeuArgLys
    ----------------------------------------------------
```

```
2101  gctccacctc atcactcctg gaagcctgca ctttccgtag accttaa
665   CysSerThr SerSerLeuLeu GluAlaCys ThrPheArg ArgPro---
      ----------------------------------------------------↑End
```

I.2. The hSTf model PDB output file

Due to the large volume of text associated with this file, it has been decided not to include it in this thesis. However, this file can be provided on request.

I.3. Modeller Log File

This file was created following the creation of the hSTf model using Modeller 6v2.

```
PROTEIN STRUCTURE MODELLING BY SATISFACTION OF SPATIAL RESTRAINTS
                Copyright(c) 1989-2002 Andrej Sali
                     All Rights Reserved
Kind, OS, HostName, Kernel, Processor: 4, IRIX evanssg1 6.5 IP22
   Date and time of compilation          : 07/05/2002 17:12:26
Job starting time (YY/MM/DD HH:MM:SS): 2004/08/25  12:03:19.255
```

```
check_ali___> Checking the sequence-structure alignment.
Implied target CA(i)-CA(i+1) distances longer than  8.0 angstroms:
ALN_POS  TMPL  RID1  RID2  NAM1  NAM2     DIST
-------------------------------------------
END OF TABLE
patch_s_522_> Number of disulfides patched in MODEL:       19
delete__443_> Restraints marked for deletion were removed.
              Total number of restraints before, now:    72348    67436

>> ENERGY; Differences between the model's features and restraints:
ID1, ID2                                          :     9999       1
Number of all residues in MODEL                   :      677
Number of all, selected real atoms                :     5252    5252
Number of all, selected pseudo atoms              :        0       0
Number of all static, selected restraints         :    67436   67436
COVALENT_CYS                                      :        F
NONBONDED_SEL_ATOMS                               :        1
Number of non-bonded pairs (excluding 1-2,1-3,1-4):    11393
Dynamic pairs routine                             : 2, NATM x NATM cell sorting
Atomic shift for contacts update (UPDATE_DYNAMIC) :    0.390
LENNARD_JONES_SWITCH                              :    6.500   7.500
COULOMB_JONES_SWITCH                              :    6.500   7.500
RESIDUE_SPAN_RANGE                                :        0     800
NLOGN_USE                                         :       15
CONTACT_SHELL                                     :    4.000
DYNAMIC_PAIRS,_SPHERE,_COULOMB,_LENNARD,_MODELLER :        T       T       F       F       F
SCHEDULE_STEP, N_SCHEDULE                         :       13      13
SPHERE_STDV                                       :    0.050
RADII_FACTOR                                      :    0.820
Current energy                                    :        4098.1191
Symmetry term within energy                       :           0.0000

Summary of the restraint violations:
              NUM      ... number of restraints.
              NUMVI    ... number of restraints with RVIOL > VIOL_REPORT_CUT[i].
              RVIOL    ... relative difference from the best value.
              NUMVP    ... number of restraints with -Ln(pdf) > VIOL_REPORT_CUT2[i].
              RMS_1    ... RMS(feature, minimally_violated_basis_restraint, NUMB).
              RMS_2    ... RMS(feature, best_value, NUMB).
              MOL.PDF  ... scaled contribution to -Ln(Molecular pdf).

#                    RESTRAINT_GROUP      NUM   NUMVI  NUMVP   RMS_1   RMS_2        MOL.PDF      S_i
-----------------------------------------------------------------------------------------------------
1 Bond length potential                :  5371      0      0   0.005   0.005       42.525       1.000
2 Bond angle potential                 :  7260      7     36   2.689   2.689       769.50       1.000
3 Stereochemical cosine torsion poten: :  3343      0    105  48.160  48.160       1217.9       1.000
4 Stereochemical improper torsion pot: :  2213      0      3   1.426   1.426       106.70       1.000
5 Soft-sphere overlap restraints       : 11393      0      0   0.002   0.002       5.1942       1.000
6 Lennard-Jones 6-12 potential         :     0      0      0   0.000   0.000       0.0000       1.000
7 Coulomb point-point electrostatic p: :     0      0      0   0.000   0.000       0.0000       1.000
8 H-bonding potential                  :     0      0      0   0.000   0.000       0.0000       1.000
9 Distance restraints 1 (CA-CA)        : 15570      0      0   0.096   0.096       165.64       1.000
10 Distance restraints 2 (N-O)         : 15650      0      4   0.151   0.151       397.38       1.000
11 Mainchain Phi dihedral restraints   :     0      0      0   0.000   0.000       0.0000       1.000
12 Mainchain Psi dihedral restraints   :     0      0      0   0.000   0.000       0.0000       1.000
13 Mainchain Omega dihedral restraints:   676      0      5   4.478   4.478       158.32       1.000
14 Sidechain Chi_1 dihedral restraints:   570      0      7  57.713  57.713       94.010       1.000
15 Sidechain Chi_2 dihedral restraints:   455      0      7  68.652  68.652       171.53       1.000
16 Sidechain Chi_3 dihedral restraints:   202      0      1  65.499  65.499       105.58       1.000
17 Sidechain Chi_4 dihedral restraints:    84      0      0  73.123  73.123       47.956       1.000
18 Disulfide distance restraints       :    19      0      0   0.015   0.015      0.73262       1.000
19 Disulfide angle restraints          :    38      0      0   2.247   2.247       4.2370       1.000
20 Disulfide dihedral angle restraints:     0      0      0   0.000   0.000       0.0000       1.000
21 Lower bound distance restraints     :     0      0      0   0.000   0.000       0.0000       1.000
22 Upper bound distance restraints     :     0      0      0   0.000   0.000       0.0000       1.000
23 Distance restraints 3 (SDCH-MNCH)   :  9645      0      0   0.345   0.345       227.20       1.000
24 Sidechain Chi_5 dihedral restraints:     0      0      0   0.000   0.000       0.0000       1.000
25 Phi/Psi pair of dihedral restraints:   675     18     68  26.673  31.926       227.08       1.000
26 Distance restraints 4 (SDCH-SDCH)   :  5665      0      0   0.636   0.636       356.63       1.000
27 Distance restraints 5 (X-Y)         :     0      0      0   0.000   0.000       0.0000       1.000
28 NMR distance restraints 6 (X-Y)     :     0      0      0   0.000   0.000       0.0000       1.000
29 NMR distance restraints 7 (X-Y)     :     0      0      0   0.000   0.000       0.0000       1.000
30 Minimal distance restraints         :     0      0      0   0.000   0.000       0.0000       1.000
31 Non-bonded restraints               :     0      0      0   0.000   0.000       0.0000       1.000
32 Atomic accessibility restraints     :     0      0      0   0.000   0.000       0.0000       1.000
33 Atomic density restraints           :     0      0      0   0.000   0.000       0.0000       1.000

A restraint is violated when the relative difference
from the best value (RVIOL) is larger than CUTOFF.

ICSR   ... index of a restraint in the current set.
RESNO  ... residue numbers of the first two atoms.
ATM    ... IUPAC atom names of the first two atoms.
FEAT   ... the value of the feature in the model.
restr  ... the mean of the basis restraint with the smallest
           difference from the model (local minimum).
viol   ... difference from the local minimum.
rviol  ... relative difference from the local minimum.
RESTR  ... the best value (global minimum).
VIOL   ... difference from the best value.
RVIOL  ... relative difference from the best value.

Feature  2                             : Bond angle potential
List of the RVIOL violations larger than   :       4.5000
    #   ICSR  RESNO1/2 ATM1/2   INDATM1/2    FEAT   restr    viol   rviol   RESTR    VIOL   RVIOL
    1  11308 553N 553N ND2 CG   4250 4248  139.04  122.50   16.54    4.60  122.50   16.54    4.60
Feature 25                             : Phi/Psi pair of dihedral restraints
List of the RVIOL violations larger than   :       6.5000

    #   ICSR  RESNO1/2 ATM1/2   INDATM1/2    FEAT   restr    viol   rviol   RESTR    VIOL   RVIOL
    1  18253   9V  10S C    N      73   75   82.60 -136.60  144.17    4.97  -64.10 -155.24   16.77
```

1		10S	10S	N	CA	75	76	-177.86	151.20			-35.00			
2	18369	125G	126W	C	N	950	952	-108.61	-63.00	46.85	8.18	-63.00	46.85	8.18	
2		126W	126W	N	CA	952	953	-54.89	-44.20			-44.20			
3	18396	152F	153S	C	N	1172	1174	-105.85	-136.60	93.52	4.00	-64.10	106.41	6.62	
3		153S	153S	N	CA	1174	1175	62.87	151.20			-35.00			
4	18574	330P	331E	C	N	2556	2558	56.68	-69.30	134.21	11.61	-69.30	134.21	11.61	
4		331E	331E	N	CA	2558	2559	-171.22	142.50			142.50			
5	18578	334T	335D	C	N	2584	2586	-100.67	-96.50	8.30	0.35	54.50	174.80	14.87	
5		335D	335D	N	CA	2586	2587	121.38	114.20			40.90			
6	18580	336E	337C	C	N	2601	2603	-67.80	-63.00	72.38	8.51	-63.00	72.38	8.51	
6		337C	337C	N	CA	2603	2604	31.12	-41.10			-41.10			
7	18581	337C	338K	C	N	2607	2609	-69.15	-62.90	83.38	10.50	-62.90	83.38	10.50	
7		338K	338K	N	CA	2609	2610	42.34	-40.80			-40.80			
8	18582	338K	339P	C	N	2616	2618	71.31	-64.50	135.86	13.71	-64.50	135.86	13.71	
8		339P	339P	N	CA	2618	2619	143.28	147.20			147.20			
9	18589	345L	346S	C	N	2672	2674	47.74	-72.40	124.03	9.59	-64.10	-179.42	13.23	
9		346S	346S	N	CA	2674	2675	-176.77	152.40			-35.00			
10	18604	360S	361V	C	N	2799	2801	64.89	55.90	10.95	1.34	-125.40	-164.25	10.38	
10		361V	361V	N	CA	2801	2802	45.75	39.50			143.30			
11	18624	380M	381N	C	N	2944	2946	-104.23	-63.20	83.58	8.99	-63.20	83.58	8.99	
11		381N	381N	N	CA	2946	2947	31.71	-41.10			-41.10			
12	18658	414D	415N	C	N	3189	3191	60.51	55.90	28.25	1.65	-63.20	134.47	21.99	
12		415N	415N	N	CA	3191	3192	11.63	39.50			-41.10			
13	18684	440D	441N	C	N	3386	3388	-107.78	-63.20	85.41	9.18	-63.20	85.41	9.18	
13		441N	441N	N	CA	3388	3389	31.75	-41.10			-41.10			
14	18701	457G	458W	C	N	3505	3507	-111.26	-63.00	50.49	8.88	-63.00	50.49	8.88	
14		458W	458W	N	CA	3507	3508	-59.04	-44.20			-44.20			
15	18742	498G	499S	C	N	3826	3828	-31.82	-136.60	105.11	4.39	-64.10	168.57	10.72	
15		499S	499S	N	CA	3828	3829	159.55	151.20			-35.00			
16	18743	499S	500G	C	N	3832	3834	-58.69	-62.40	9.68	1.64	82.20	146.67	11.02	
16		500G	500G	N	CA	3834	3835	-32.27	-41.20			8.50			
17	18782	538Q	539N	C	N	4137	4139	-123.82	-119.90	94.27	4.37	-63.20	103.52	11.18	
17		539N	539N	N	CA	4139	4140	42.81	137.00			-41.10			
18	18794	550K	551N	C	N	4227	4229	-106.32	-63.20	85.54	9.19	-63.20	85.54	9.19	
18		551N	551N	N	CA	4229	4230	32.78	-41.10			-41.10			

report_____> Distribution of short non-bonded contacts:

DISTANCE1:	0.00	2.10	2.20	2.30	2.40	2.50	2.60	2.70	2.80	2.90	3.00	3.10	3.20	3.30	3.40
DISTANCE2:	2.10	2.20	2.30	2.40	2.50	2.60	2.70	2.80	2.90	3.00	3.10	3.20	3.30	3.40	3.50
FREQUENCY:	0	0	0	0	0	25	83	257	360	461	469	659	746	851	891

<< end of ENERGY.

Dynamically allocated memory at	finish [B,kB,MB]:	29583587	28890.223 28.213
Starting time	:	2004/08/25	12:03:19.255
Closing time	:	2004/08/25	13:12:32.963
Total CPU time [seconds]	:	0.00	

I.4. Human serum transferrin model bump table

A table of atomic pairs that bump according to WHATIF analysis of the hSTf model.

Key. A: Bump Number. B + C: Residue One. D: Atom. E + F: Residue Two. G: Atom. H: How much shorter the contact is than the accepted limited (Å). I: Distance (Å) between the two centres of the atoms. J: Status of the atomic pair. K: Explanation of bump (BF: B-factor > 80, HB: Hydrogen Bonding).

A	B	C	D	E	F	G	H	I	J	K
1	1	ASP	CG	2	LYS	N	0.284	2.816	INTRA	BF
2	1	ASP	CB	2	LYS	N	0.023	2.677	INTRA	BF
3	2	LYS	CB	260	LEU	CD2	0.048	3.152	INTRA	BF
4	4	VAL	CG1	261	ILE	CD1	0.04	3.16	INTRA	BF
5	4	VAL	CG2	264	LEU	CD2	0.027	3.173	INTRA	BF
6	5	ARG	NH2	56	ASP	OD1	0.036	2.664	INTRA	BF
7	5	ARG	NH1	257	LYS	NZ	0.026	2.974	INTRA	BF
8	7	CYS	SG	58	VAL	CB	0.37	3.03	INTRA	BF
9	7	CYS	SG	58	VAL	CG2	0.363	3.037	INTRA	BF
10	7	CYS	CA	38	VAL	CG2	0.015	3.185	INTRA	BF
11	10	SER	CB	178	SER	OG	0.019	2.781	INTRA	BF
12	11	GLU	O	12	HIS	C	0.011	2.789	INTRA	BF
13	15	THR	CG2	287	HIS	NE2	0.208	2.892	INTRA	BF
14	15	THR	CG2	287	HIS	CE1	0.172	3.028	INTRA	BF
15	15	THR	O	18	GLN	CG	0.06	2.74	INTRA	BF
16	17	CYS	C	37	CYS	SG	0.007	3.393	INTRA	BF
17	18	GLN	CD	19	SER	N	0.274	2.826	INTRA	BF
18	18	GLN	CB	21	ARG	NH1	0.233	2.867	INTRA	BF
19	18	GLN	CG	19	SER	N	0.099	3.001	INTRA	BF
20	18	GLN	O	21	ARG	CG	0.053	2.747	INTRA	BF
21	21	ARG	CG	22	ASP	N	0.272	2.828	INTRA	BF
22	21	ARG	CB	35	VAL	CG2	0.263	2.937	INTRA	BF
23	21	ARG	CA	35	VAL	CG2	0.184	3.016	INTRA	BF
24	23	HIS	CD2	280	PHE	CB	0.202	2.998	INTRA	BF
25	23	HIS	CE1	280	PHE	CB	0.169	3.031	INTRA	BF
26	23	HIS	CD2	278	LYS	NZ	0.149	2.951	INTRA	BF
27	23	HIS	NE2	280	PHE	CB	0.148	2.952	INTRA	BF
28	23	HIS	CG	280	PHE	CB	0.046	3.154	INTRA	BF
29	24	MET	CE	272	PHE	CE2	0.25	2.95	INTRA	BF
30	24	MET	O	25	LYS	C	0.062	2.738	INTRA	BF
31	28	ILE	CG2	29	PRO	N	0.319	2.781	INTRA	BF
32	28	ILE	CG2	32	GLY	CA	0.259	2.941	INTRA	BF
33	28	ILE	CG2	29	PRO	CD	0.169	3.031	INTRA	BF
34	28	ILE	CA	29	PRO	CD	0.024	2.776	INTRA	BF
35	31	ASP	O	32	GLY	C	0.108	2.692	INTRA	BF
36	32	GLY	O	33	PRO	C	0.035	2.765	INTRA	BF
37	32	GLY	CA	33	PRO	CD	0.02	2.78	INTRA	BF
38	33	PRO	CB	264	LEU	CD2	0.167	3.033	INTRA	BF
39	42	SER	CB	177	CYS	SG	0.125	3.275	INTRA	BF
40	43	TYR	CG	44	LEU	N	0.13	2.97	INTRA	BF
41	50	ILE	CD1	56	ASP	C	0.058	3.142	INTRA	BF
42	52	ALA	O	53	ASN	CB	0.065	2.735	INTRA	BF
43	53	ASN	OD1	252	ARG	CZ	0.056	2.744	INTRA	BF
44	61	ASP	CG	62	ALA	N	0.11	2.99	INTRA	BF
45	65	VAL	CG1	77	PRO	CG	0.04	3.16	INTRA	BF
46	66	TYR	CE1	318	VAL	CG2	0.2	3	INTRA	BF
47	69	TYR	CB	309	ALA	CB	0.222	2.978	INTRA	BF
48	69	TYR	CD1	309	ALA	CB	0.18	3.02	INTRA	BF
49	69	TYR	CG	309	ALA	CB	0.122	3.078	INTRA	BF
50	71	ALA	CA	74	ASN	ND2	0.237	2.863	INTRA	BF
51	73	ASN	O	74	ASN	C	0.023	2.777	INTRA	BF
52	78	VAL	CG2	79	VAL	N	0 129	2.971	INTRA	BF
53	78	VAL	CG1	258	GLU	CG	0.124	3.076	INTRA	BF
54	79	VAL	CG1	80	ALA	N	0.188	2.912	INTRA	BF
55	82	PHE	CE2	299	GLY	C	0.255	2.945	INTRA	BF
56	82	PHE	CE1	298	HIS	CB	0.251	2.949	INTRA	BF
57	82	PHE	CE2	299	GLY	CA	0.206	2 994	INTRA	BF

58	82	PHE	CZ	299	GLY	CA	0.085	3.115	INTRA	BF
59	82	PHE	CD2	299	GLY	C	0.038	3.162	INTRA	BF
60	82	PHE	CE2	300	PHE	N	0.008	3.092	INTRA	BF
61	84	GLY	N	298	HIS	CD2	0.226	2.874	INTRA	BF
62	84	GLY	C	298	HIS	NE2	0.215	2.885	INTRA	BF
63	84	GLY	C	298	HIS	CD2	0.091	3.109	INTRA	BF
64	84	GLY	O	298	HIS	NE2	0.08	2.62	INTRA	BF
65	85	SER	OG	87	GLU	CD	0.015	2.785	INTRA	BF
66	87	GLU	C	89	PRO	CD	0.04	3.16	INTRA	BF
67	88	ASP	N	89	PRO	CD	0.037	2.963	INTRA	BF
68	88	ASP	CA	89	PRO	CD	0.01	2.79	INTRA	BF
69	92	PHE	CE1	677	PRO	O	0.044	2.756	INTRA	BF
70	95	ALA	CB	128	ILE	CD1	0.253	2.947	INTRA	BF
71	98	VAL	CG1	99	VAL	N	0.062	3.038	INTRA	BF
72	99	VAL	CG2	100	LYS	N	0.282	2.818	INTRA	BF
73	99	VAL	CG2	230	ARG	NH2	0.192	2.908	INTRA	BF
74	99	VAL	CG1	230	ARG	NH2	0.01	3.09	INTRA	BF
75	101	LYS	O	102	ASP	CG	0.064	2.736	INTRA	BF
76	103	SER	OG	105	PHE	CE2	0.019	2.781	INTRA	BF
77	105	PHE	CB	109	GLN	NE2	0.232	2.868	INTRA	BF
78	105	PHE	CA	109	GLN	CD	0.141	3.059	INTRA	BF
79	105	PHE	CB	109	GLN	CD	0.02	3.18	INTRA	BF
80	105	PHE	CA	109	GLN	NE2	0.014	3.086	INTRA	BF
81	106	GLN	N	109	GLN	OE1	0.078	2.622	INTRA	BF
82	106	GLN	OE1	228	ASN	C	0.019	2.781	INTRA	BF
83	110	LEU	O	111	ARG	C	0.082	2.718	INTRA	BF
84	111	ARG	CG	151	PHE	CE1	0.258	2.942	INTRA	BF
85	111	ARG	CG	151	PHE	CZ	0.246	2.954	INTRA	BF
86	114	LYS	CE	153	SER	O	0.026	2.774	INTRA	BF
87	116	CYS	SG	171	LEU	CD1	0.206	3.194	INTRA	BF
88	116	CYS	SG	156	CYS	CB	0.116	3.284	INTRA	BF
89	123	SER	CB	317	TYR	OH	0.021	2.779	INTRA	BF
90	126	TRP	CZ2	144	LEU	CD1	0.251	2.949	INTRA	BF
91	128	ILE	CD1	244	VAL	CG1	0.25	2.95	INTRA	BF
92	128	ILE	N	129	PRO	CD	0.107	2.893	INTRA	BF
93	128	ILE	CB	129	PRO	CD	0.056	3.144	INTRA	BF
94	133	LEU	O	136	ASP	CG	0.059	2.741	INTRA	BF
95	134	TYR	CD1	329	CYS	SG	0.369	3.031	INTRA	BF
96	134	TYR	CE2	144	LEU	CD2	0.122	3.078	INTRA	BF
97	134	TYR	OH	141	ARG	CB	0.018	2.782	INTRA	BF
98	134	TYR	CD2	144	LEU	CD2	0.013	3.187	INTRA	BF
99	134	TYR	O	135	CYS	C	0.01	2.79	INTRA	BF
100	135	CYS	SG	330	PRO	CD	0.025	3.375	INTRA	BF
101	137	LEU	CB	138	PRO	CD	0.069	3.131	INTRA	BF
102	137	LEU	CD2	151	PHE	CD1	0.057	3.143	INTRA	BF
103	137	LEU	O	138	PRO	C	0.054	2.746	INTRA	BF
104	137	LEU	CA	138	PRO	CD	0.024	2.776	INTRA	BF
105	138	PRO	O	139	GLU	C	0.052	2.748	INTRA	BF
106	139	GLU	CB	140	PRO	CD	0.046	3.154	INTRA	BF
107	142	LYS	CA	143	PRO	C	0.088	3.112	INTRA	BF
108	142	LYS	CA	143	PRO	CA	0.075	2.725	INTRA	BF
109	145	GLU	CB	165	PHE	CE2	0.249	2.951	INTRA	BF
110	145	GLU	CB	165	PHE	CD2	0.072	3.128	INTRA	BF
111	148	VAL	CG1	149	ALA	N	0.272	2.828	INTRA	BF
112	156	CYS	SG	183	TYR	CB	0.007	3.393	INTRA	BF
113	157	ALA	O	158	PRO	C	0.032	2.768	INTRA	BF
114	161	ASP	OD1	163	THR	CG2	0.058	2.742	INTRA	BF
115	162	GLY	CA	169	CYS	SG	0.084	3.316	INTRA	BF
116	162	GLY	O	163	THR	C	0 032	2.768	INTRA	BF
117	163	THR	CG2	164	ASP	N	0.282	2.818	INTRA	BF
118	165	PHE	CB	168	LEU	CD2	0.246	2.954	INTRA	BF
119	169	CYS	O	170	GLN	C	0.053	2.747	INTRA	BF
120	169	CYS	O	172	CYS	C	0.04	2.76	INTRA	BF
121	171	LEU	C	172	CYS	SG	0.093	3.307	INTRA	BF
122	172	CYS	SG	183	TYR	CB	0.422	2.978	INTRA	BF
123	172	CYS	O	173	PRO	C	0.057	2.743	INTRA	BF
124	172	CYS	SG	183	TYR	CG	0.051	3.349	INTRA	BF
125	174	GLY	O	175	CYS	C	0.053	2.747	INTRA	BF
126	176	GLY	C	181	ASN	ND2	0.239	2.861	INTRA	BF
127	178	SER	O	181	ASN	ND2	0.045	2.655	INTRA	BF
128	179	THR	N	180	LEU	N	0.009	2.591	INTRA	BF
129	183	TYR	C	188	GLY	CA	0.19	3.01	INTRA	BF
130	184	PHE	CG	185	GLY	N	0.138	2.962	INTRA	BF
131	190	PHE	CE1	208	ILE	CG2	0.121	3.079	INTRA	BF
132	190	PHE	CZ	212	LEU	CD1	0.097	3.103	INTRA	BF
133	193	LEU	CD1	221	TYR	CE2	0.254	2.946	INTRA	BF
134	194	LYS	CD	221	TYR	OH	0.018	2.782	INTRA	BF
135	200	VAL	CG2	201	ALA	N	0.144	2.956	INTRA	BF
136	204	LYS	NZ	294	LYS	NZ	0.135	2.865	INTRA	BF
137	205	HIS	CG	206	SER	N	0.037	3.063	INTRA	BF
138	206	SER	O	207	THR	C	0.063	2.737	INTRA	BF
139	207	THR	O	208	ILE	C	0.031	2.769	INTRA	BF
140	208	ILE	CD1	233	VAL	CG1	0.258	2.942	INTRA	BF
141	208	ILE	CD1	218	ARG	CG	0.249	2.951	INTRA	BF
142	208	ILE	CD1	233	VAL	CG2	0.239	2.961	INTRA	BF
143	209	PHE	CE1	218	ARG	NE	0.107	2.993	INTRA	BF
144	223	LEU	CD2	233	VAL	CA	0.248	2.952	INTRA	BF
145	227	ASP	O	228	ASN	CB	0.055	2.745	INTRA	BF
146	229	THR	CG2	230	ARG	N	0.274	2.826	INTRA	BF
147	231	LYS	CD	239	CYS	SG	0.369	3.031	INTRA	BF
148	231	LYS	CB	232	PRO	CD	0.096	3.104	INTRA	BF
149	231	LYS	CA	232	PRO	CD	0.053	2.747	INTRA	BF
150	235	GLU	O	236	TYR	C	0.022	2.778	INTRA	BF
151	236	TYR	CD1	237	LYS	N	0.098	3.002	INTRA	BF
152	236	TYR	O	237	LYS	C	0.059	2.741	INTRA	BF
153	237	LYS	NZ	238	ASP	CG	0.189	2.911	INTRA	BF
154	237	LYS	CD	238	ASP	N	0.182	2.918	INTRA	BF
155	237	LYS	CD	238	ASP	OD1	0.055	2.745	INTRA	BF
156	244	VAL	CG1	317	TYR	CE1	0.101	3.099	INTRA	BF
157	244	VAL	CA	245	PRO	CD	0.028	2.772	INTRA	BF
158	247	HIS	CD2	294	LYS	CD	0.136	3.064	INTRA	BF
159	247	HIS	C	248	THR	CG2	0.086	3.114	INTRA	BF
160	248	THR	CG2	312	TYR	CZ	0.112	3.088	INTRA	BF
161	248	THR	CG2	312	TYR	OH	0.017	2.783	INTRA	BF
162	259	ASP	CG	260	LEU	N	0.225	2.875	INTRA	BF
163	259	ASP	CG	260	LEU	CD1	0.135	3.065	INTRA	BF
164	259	ASP	OD1	260	LEU	CD1	0 054	2.746	INTRA	BF
165	260	LEU	O	264	LEU	CD2	0.047	2.753	INTRA	BF
166	260	LEU	O	261	ILE	C	0.023	2.777	INTRA	BF
167	266	ASN	O	269	GLN	CG	0.011	2.789	INTRA	BF
168	267	GLN	O	268	ALA	C	0.016	2.784	INTRA	
169	269	GLN	O	270	GLU	C	0.055	2.745	INTRA	BF
170	275	ASP	CG	276	LYS	N	0.237	2.863	INTRA	BF
171	276	LYS	CG	277	SER	N	0.098	3.002	INTRA	BF
172	278	LYS	CG	279	GLU	N	0.273	2.827	INTRA	BF
173	291	LEU	C	292	LEU	CD1	0.006	3.194	INTRA	BF
174	303	VAL	CA	304	PRO	CD	0.056	2.744	INTRA	BF
175	303	VAL	CG1	304	PRO	N	0.028	3.072	INTRA	BF
176	304	PRO	CA	305	PRO	CD	0.093	2.707	INTRA	BF
177	306	ARG	CD	668	LEU	CA	0.083	3.117	INTRA	BF
178	306	ARG	NH1	311	MET	CE	0.078	3.022	INTRA	BF
179	307	MET	CE	671	ALA	O	0.059	2.741	INTRA	BF
180	309	ALA	O	313	LEU	CD1	0.059	2.741	INTRA	BF
181	309	ALA	O	310	LYS	C	0.013	2.787	INTRA	BF
182	310	LYS	CE	381	ASN	ND2	0.234	2.866	INTRA	BF
183	310	LYS	NZ	381	ASN	CG	0.173	2.927	INTRA	BF
184	310	LYS	CE	381	ASN	CG	0.158	3.042	INTRA	BF
185	310	LYS	CD	381	ASN	CB	0.058	3.142	INTRA	BF
186	311	MET	SD	671	ALA	CB	0.279	3.121	INTRA	BF
187	316	GLU	O	319	THR	CG2	0.057	2.743	INTRA	BF
188	318	VAL	CG1	319	THR	N	0.273	2.827	INTRA	BF
189	318	VAL	O	321	ILE	CG2	0.063	2.737	INTRA	BF
190	319	THR	CG2	320	ALA	N	0.28	2.82	INTRA	BF
191	321	ILE	CG2	322	ARG	N	0.247	2.853	INTRA	BF
192	322	ARG	O	326	GLU	CG	0.053	2.747	INTRA	BF
193	327	GLY	N	328	THR	N	0.033	2.567	INTRA	BF
194	329	CYS	CA	330	PRO	CD	0.093	2.707	INTRA	BF
195	329	CYS	O	330	PRO	C	0.057	2.743	INTRA	BF
196	330	PRO	C	331	GLU	O	0.085	2.715	INTRA	BF
197	330	PRO	O	331	GLU	C	0.022	2.778	INTRA	BF
198	330	PRO	CG	331	GLU	N	0.02	3.08	INTRA	BF
199	331	GLU	O	332	ALA	C	0.076	2.724	INTRA	BF
200	331	GLU	C	332	ALA	C	0.008	2.792	INTRA	BF
201	332	ALA	CA	333	PRO	CD	0.11	2.69	INTRA	BF
202	334	THR	CG2	335	ASP	N	0.272	2.828	INTRA	BF
203	336	GLU	CG	594	CYS	SG	0.249	3.151	INTRA	BF
204	336	GLU	O	338	LYS	N	0.083	2.617	INTRA	BF
205	338	LYS	C	339	PRO	CD	0.125	2.175	INTRA	BF
206	339	PRO	CB	363	LYS	C	0.009	3.191	INTRA	BF
207	340	VAL	CG1	341	LYS	N	0.117	2.983	INTRA	BF
208	341	LYS	CD	384	ALA	CA	0.249	2.951	INTRA	BF
209	341	LYS	CG	385	ASP	N	0.094	3.006	INTRA	BF
210	341	LYS	CD	383	GLU	O	0.062	2.738	INTRA	BF
211	343	CYS	SG	387	MET	SD	0.449	3.151	INTRA	BF
212	345	LEU	CD1	387	MET	CE	0.254	2.946	INTRA	BF
213	347	HIS	CE1	351	LEU	CD1	0.167	3.033	INTRA	BF
214	349	GLU	CG	627	LEU	CD1	0.161	3.039	INTRA	BF
215	351	LEU	O	352	LYS	C	0.001	2.799	INTRA	
216	353	CYS	C	366	CYS	SG	0.188	3.212	INTRA	BF
217	353	CYS	O	354	ASP	C	0.031	2.769	INTRA	BF
218	355	GLU	CG	617	PHE	CD1	0.012	3.188	INTRA	BF
219	357	SER	O	358	VAL	C	0.066	2.734	INTRA	BF
220	360	SER	O	361	VAL	CG2	0.083	2.717	INTRA	BF
221	367	VAL	CG1	368	SER	N	0.159	2.941	INTRA	BF
222	369	ALA	CB	378	LYS	CE	0.025	3.175	INTRA	BF
223	372	THR	CG2	373	GLU	N	0.278	2.822	INTRA	BF
224	372	THR	O	376	ILE	CD1	0.064	2.736	INTRA	BF
225	372	THR	CG2	393	PHE	CE1	0.062	3.138	INTRA	BF
226	375	CYS	SG	387	MET	SD	0.394	3.206	INTRA	BF
227	375	CYS	CB	387	MET	SD	0.154	3.246	INTRA	BF
228	377	ALA	CA	380	MET	CE	0.252	2.948	INTRA	BF
229	379	ILE	O	588	ARG	CD	0.069	2.731	INTRA	BF
230	380	MET	SD	400	CYS	CB	0.159	3.241	INTRA	BF
231	381	ASN	OD1	383	GLU	CG	0.055	2.745	INTRA	BF
232	390	ASP	CG	391	GLY	N	0.229	2.871	INTRA	BF
233	391	GLY	O	394	VAL	CG1	0.065	2.735	INTRA	BF
234	394	VAL	CG1	395	TYR	N	0.174	2.926	INTRA	BF
235	394	VAL	CG1	649	LEU	CD1	0.124	3.076	INTRA	BF
236	398	GLY	O	399	LYS	C	0.02	2.78	INTRA	BF
237	399	LYS	N	400	CYS	N	0.019	2.581	INTRA	BF
238	403	VAL	CG1	589	LYS	CD	0.237	2.963	INTRA	BF
239	405	VAL	CG1	406	LEU	N	0.283	2.817	INTRA	BF
240	409	ASN	CB	415	ASN	ND2	0.005	3.095	INTRA	BF
241	410	TYR	CE2	582	ASN	CG	0.064	3.136	INTRA	BF
242	411	ASN	CB	421	GLU	OE2	0.054	2.746	INTRA	BF
243	412	LYS	CD	414	ASP	N	0.28	2.82	INTRA	BF
244	412	LYS	CD	413	SER	N	0.278	2.822	INTRA	BF
245	412	LYS	O	634	VAL	CG1	0.06	2.74	INTRA	BF
246	412	LYS	CB	415	ASN	OD1	0.058	2.742	INTRA	BF
247	413	SER	CA	634	VAL	CG1	0.255	2.945	INTRA	BF
248	413	SER	N	414	ASP	N	0.083	2.517	INTRA	BF
249	413	SER	C	414	ASP	O	0.075	2.725	INTRA	BF
250	414	ASP	CG	415	ASN	N	0.273	2.827	INTRA	BF
251	414	ASP	C	416	CYS	N	0.139	2.761	INTRA	BF
252	414	ASP	O	415	ASN	CG	0.1	2.7	INTRA	BF
253	414	ASP	OD1	415	ASN	N	0.078	2.622	INTRA	BF
254	414	ASP	O	415	ASN	C	0.062	2.738	INTRA	BF
255	415	ASN	CB	418	ASP	CB	0.28	2.92	INTRA	BF
256	415	ASN	CA	418	ASP	N	0.276	2.824	INTRA	BF
257	415	ASN	C	418	ASP	N	0.265	2.835	INTRA	BF
258	415	ASN	CA	418	ASP	CB	0.199	3.001	INTRA	BF
259	415	ASN	C	416	CYS	C	0.133	2.667	INTRA	BF
260	415	ASN	O	416	CYS	C	0.069	2.731	INTRA	BF
261	415	ASN	C	417	GLU	N	0.06	2.84	INTRA	BF
262	416	CYS	N	417	GLU	N	0.182	2.418	INTRA	BF
263	419	THR	CA	420	PRO	CD	0.059	2.741	INTRA	BF
264	419	THR	CB	420	PRO	CD	0.01	3.19	INTRA	BF
265	421	GLU	CG	582	ASN	ND2	0.234	2.866	INTRA	BF
266	427	VAL	CG2	428	ALA	N	0.278	2.822	INTRA	BF
267	427	VAL	CG2	559	LEU	CD1	0.258	2 942	INTRA	BF
268	429	VAL	CG2	559	LEU	CD1	0.25	2.95	INTRA	BF
269	429	VAL	CG1	430	VAL	N	0.118	2.982	INTRA	BF
270	430	VAL	CG1	437	LEU	CD1	0.254	2.946	INTRA	BF
271	430	VAL	CG1	560	LEU	CD2	0.131	3 069	INTRA	BF

272	430	VAL	CB	437	LEU	CD1	0.07	3.13	INTRA	BF
273	432	LYS	O	433	SER	C	0.024	2.776	INTRA	BF
274	434	ALA	O	437	LEU	CD2	0.054	2.746	INTRA	BF
275	438	THR	CG2	439	TRP	N	0.294	2.806	INTRA	BF
276	438	THR	CG2	564	GLY	O	0.071	2.729	INTRA	BF
277	439	TRP	CZ2	464	LEU	CD2	0.248	2.952	INTRA	BF
278	441	ASN	CA	443	LYS	NZ	0.139	2.961	INTRA	BF
279	441	ASN	O	445	LYS	CE	0.057	2.743	INTRA	BF
280	442	LEU	O	443	LYS	C	0.075	2.725	INTRA	BF
281	443	LYS	N	477	PHE	CZ	0.046	3.054	INTRA	BF
282	448	CYS	SG	521	CYS	C	0.095	3.305	INTRA	BF
283	450	THR	OG1	454	ARG	CD	0.013	2.787	INTRA	BF
284	452	VAL	CG1	453	GLY	N	0.07	3.03	INTRA	BF
285	455	THR	CG2	580	ALA	CB	0.071	3.129	INTRA	BF
286	460	ILE	CB	461	PRO	CD	0.169	3.031	INTRA	BF
287	460	ILE	N	461	PRO	CD	0.092	2.908	INTRA	BF
288	462	MET	CE	474	PHE	CG	0.198	3.002	INTRA	BF
289	462	MET	CE	474	PHE	CB	0.123	3.077	INTRA	BF
290	464	LEU	CD2	562	LEU	CD2	0.003	3.197	INTRA	BF
291	466	TYR	O	467	ASN	C	0.06	2.74	INTRA	BF
292	471	HIS	CE1	473	ARG	CG	0.166	3.034	INTRA	BF
293	471	HIS	ND1	473	ARG	N	0.1	2.75	INTRA	BF
294	471	HIS	ND1	473	ARG	CG	0.068	3.032	INTRA	BF
295	473	ARG	CD	476	GLU	CD	0.254	2.946	INTRA	BF
296	473	ARG	CD	476	GLU	OE1	0.054	2.746	INTRA	BF
297	473	ARG	CG	476	GLU	CB	0.001	3.199	INTRA	BF
298	475	ASP	CA	492	LEU	CD2	0.265	2.935	INTRA	BF
299	475	ASP	CB	492	LEU	CD2	0.171	3.029	INTRA	BF
300	480	GLU	CD	494	LYS	CD	0.105	3.095	INTRA	BF
301	480	GLU	OE2	494	LYS	CE	0.054	2.746	INTRA	BF
302	480	GLU	OE2	494	LYS	CD	0.054	2.746	INTRA	BF
303	480	GLU	CB	495	LEU	CD1	0.024	3.176	INTRA	BF
304	481	GLY	N	495	LEU	CD1	0.28	2.82	INTRA	BF
305	481	GLY	CA	495	LEU	CD1	0.033	3.167	INTRA	BF
306	482	CYS	SG	512	TYR	CB	0.089	3.311	INTRA	BF
307	483	ALA	O	484	PRO	C	0.049	2.751	INTRA	BF
308	483	ALA	CA	484	PRO	CD	0.005	2.795	INTRA	BF
309	496	CYS	SG	504	CYS	CB	0.064	3.336	INTRA	BF
310	496	CYS	O	503	LEU	CD1	0.058	2.742	INTRA	BF
311	497	MET	O	498	GLY	C	0.09	2.71	INTRA	BF
312	499	SER	C	501	LEU	N	0.081	2.819	INTRA	BF
313	499	SER	O	500	GLY	C	0.062	2.738	INTRA	BF
314	499	SER	C	500	GLY	C	0.014	2.786	INTRA	BF
315	500	GLY	N	501	LEU	N	0.092	2.508	INTRA	BF
316	500	GLY	O	501	LEU	C	0.026	2.774	INTRA	BF
317	501	LEU	C	502	ASN	CA	0.009	2.291	INTRA	BF
318	502	ASN	CB	509	LYS	CB	0.196	3.004	INTRA	BF
319	502	ASN	OD1	509	LYS	CG	0.056	2.744	INTRA	BF
320	502	ASN	CB	505	GLU	CB	0.025	3.175	INTRA	BF
321	502	ASN	OD1	509	LYS	CB	0.013	2.787	INTRA	BF
322	505	GLU	CG	507	ASN	ND2	0.234	2.866	INTRA	BF
323	505	GLU	CG	506	PRO	CD	0.168	3.032	INTRA	BF
324	505	GLU	CA	506	PRO	CD	0.093	2.707	INTRA	BF
325	505	GLU	CD	506	PRO	CD	0.073	3.127	INTRA	BF
326	505	GLU	CG	506	PRO	N	0.068	3.032	INTRA	BF
327	508	ASN	C	510	GLU	N	0.008	2.892	INTRA	BF
328	512	TYR	CD2	521	CYS	CB	0.258	2.942	INTRA	BF
329	512	TYR	CD2	521	CYS	SG	0.151	3.249	INTRA	BF
330	512	TYR	CE2	521	CYS	CA	0.027	3.173	INTRA	BF
331	516	THR	OG1	625	LYS	CD	0.015	2.785	INTRA	BF
332	516	THR	O	517	GLY	C	0.011	2.789	INTRA	BF
333	519	PHE	CZ	557	TYR	CE2	0.261	2.939	INTRA	BF
334	519	PHE	CZ	557	TYR	CZ	0.222	2.978	INTRA	BF
335	519	PHE	CD2	539	ASN	OD1	0.056	2.744	INTRA	BF
336	520	ARG	O	521	CYS	C	0.066	2.734	INTRA	BF
337	520	ARG	O	523	VAL	CG1	0.057	2.743	INTRA	BF
338	522	LEU	O	523	VAL	C	0.02	2.78	INTRA	BF
339	523	VAL	CG1	524	GLU	N	0.281	2.819	INTRA	BF
340	528	VAL	CG2	529	ALA	N	0.069	3.031	INTRA	BF
341	530	PHE	C	531	VAL	CG1	0.061	3.139	INTRA	BF
342	532	LYS	CG	535	THR	CG2	0.195	3.005	INTRA	BF
343	533	HIS	CD2	534	GLN	N	0.233	2.867	INTRA	BF
344	535	THR	O	536	VAL	C	0.021	2.779	INTRA	BF
345	536	VAL	CB	537	PRO	CD	0.174	3.026	INTRA	BF
346	536	VAL	N	537	PRO	CD	0.092	2.908	INTRA	BF
347	537	PRO	O	543	LYS	CD	0.058	2.742	INTRA	BF
348	539	ASN	CB	544	ASN	CB	0.057	3.143	INTRA	BF
349	542	GLY	N	543	LYS	N	0.008	2.592	INTRA	B3
350	544	ASN	ND2	625	LYS	NZ	0.174	2.826	INTRA	BF
351	544	ASN	CA	545	PRO	CD	0.101	2.699	INTRA	BF
352	553	ASN	O	556	ASP	CG	0 04	2.76	INTRA	BF
353	556	ASP	OD1	557	TYR	CE2	0.084	2.716	INTRA	BF
354	556	ASP	OD1	557	TYR	CD2	0.061	2.739	INTRA	BF
355	559	LEU	CD2	569	VAL	CA	0.258	2.942	INTRA	BF
356	561	CYS	SG	565	THR	CG2	0.132	3.268	INTRA	BF
357	561	CYS	CB	565	THR	CG2	0.035	3.165	INTRA	BF
358	561	CYS	N	575	CYS	SG	0.004	3.296	INTRA	BF
359	564	GLY	N	565	THR	N	0.094	2.506	INTRA	BF
360	567	LYS	CA	568	PRO	CD	0.038	2.762	INTRA	BF
361	568	PRO	O	571	GLU	CB	0.017	2.783	INTRA	BF
362	569	VAL	CG1	570	GLU	N	0.109	2.991	INTRA	BF
363	571	GLU	O	572	TYR	C	0.078	2.722	INTRA	BF
364	572	TYR	CD1	573	ALA	N	0.276	2.824	INTRA	BF
365	572	TYR	CG	573	ALA	N	0.085	3.015	INTRA	BF
366	572	TYR	O	573	ALA	C	0.074	2.726	INTRA	BF
367	585	VAL	CG1	586	VAL	N	0.082	3.018	INTRA	BF
368	587	THR	CG2	588	ARG	N	0.01	3.09	INTRA	BF
369	587	THR	CG2	592	GLU	CA	0	3.2	INTRA	BF
370	588	ARG	NH1	591	LYS	CB	0.234	2.866	INTRA	BF
371	589	LYS	O	590	ASP	C	0.064	2.736	INTRA	BF
372	590	ASP	CG	591	LYS	N	0.232	2.868	INTRA	BF
373	591	LYS	O	592	GLU	C	0.066	2.734	INTRA	BF
374	593	ALA	CA	596	HIS	CD2	0.21	2.99	INTRA	BF
375	593	ALA	O	594	CYS	C	0.071	2.729	INTRA	BF
376	593	ALA	O	596	HIS	CD2	0.018	2.785	INTRA	BF
377	595	VAL	O	598	ILE	CG2	0.04	2.76	INTRA	BF
378	596	HIS	ND1	597	LYS	N	0.209	2.791	INTRA	BF
379	596	HIS	CE1	597	LYS	CG	0.17	3.03	INTRA	BF
380	596	HIS	CG	597	LYS	N	0.154	2.946	INTRA	BF
381	596	HIS	O	597	LYS	C	0.063	2.737	INTRA	BF
382	598	ILE	CG2	599	LEU	N	0.275	2.825	INTRA	BF
383	599	LEU	O	602	GLN	CG	0.055	2.745	INTRA	BF
384	599	LEU	O	600	ARG	C	0.055	2.745	INTRA	BF
385	602	GLN	CD	603	GLN	N	0.17	2.93	INTRA	BF
386	602	GLN	NE2	606	PHE	CB	0.082	3.018	INTRA	BF
387	602	GLN	CG	603	GLN	N	0.05	3.05	INTRA	BF
388	603	GLN	O	604	HIS	C	0.047	2.753	INTRA	BF
389	609	ASN	CG	610	VAL	N	0.138	2.962	INTRA	BF
390	609	ASN	O	610	VAL	C	0.062	2.738	INTRA	BF
391	610	VAL	CG1	613	CYS	SG	0.127	3.273	INTRA	BF
392	619	LEU	CD2	629	PHE	CD1	0.268	2.932	INTRA	BF
393	619	LEU	CD2	629	PHE	CE1	0.143	3.057	INTRA	BF
394	619	LEU	CD1	629	PHE	CE1	0.102	3.098	INTRA	BF
395	623	GLU	CG	624	THR	N	0.262	2.838	INTRA	BF
396	624	THR	CB	625	LYS	N	0.05	2.65	INTRA	BF
397	624	THR	OG1	627	LEU	CD2	0.032	2.768	INTRA	BF
398	627	LEU	C	629	PHE	N	0.007	2.893	INTRA	BF
399	638	LYS	CD	639	LEU	N	0.28	2.82	INTRA	BF
400	638	LYS	CD	640	HIS	N	0.23	2.87	INTRA	BF
401	639	LEU	O	640	HIS	C	0.037	2.763	INTRA	BF
402	644	THR	CG2	645	TYR	N	0.213	2.887	INTRA	BF
403	644	THR	CB	647	LYS	CG	0.025	3.175	INTRA	BF
404	645	TYR	CG	646	GLU	N	0.159	2.941	INTRA	BF
405	645	TYR	O	649	LEU	CD2	0.063	2.737	INTRA	BF
406	651	GLU	O	654	VAL	CG1	0.065	2.735	INTRA	BF
407	651	GLU	O	652	GLU	C	0.056	2.744	INTRA	BF
408	654	VAL	CG1	655	LYS	N	0.273	2.827	INTRA	BF
409	654	VAL	O	655	LYS	C	0.04	2.76	INTRA	BF
410	658	GLY	O	659	ASN	C	0.007	2.793	INTRA	
411	659	ASN	O	662	LYS	CD	0.052	2.748	INTRA	BF
412	660	LEU	O	661	ARG	C	0.055	2.745	INTRA	BF
413	661	ARG	O	662	LYS	C	0.032	2.768	INTRA	BF
414	663	CYS	O	664	SER	C	0.061	2.739	INTRA	BF
415	667	SER	O	670	GLU	CG	0.054	2.746	INTRA	
416	670	GLU	CG	671	ALA	N	0.1	3	INTRA	BF
417	670	GLU	O	673	THR	CG2	0.07	2.73	INTRA	BF
418	674	PHE	O	675	ARG	C	0.006	2.794	INTRA	BF

I.5. Rabbit serum transferrin bump table

A table of atomic pairs that bump according to WHATIF analysis of rSTf. Key. A: Bump Number. B + C: Residue One. D: Atom. E + F: Residue Two. G: Atom. H: How much shorter the contact is than the accepted limited (Å). I: Distance (Å)

between the two centres of the atoms. J: Status of the atomic pair. K: Explanation of bump (BF: B-factor > 80, HB: Hydrogen Bonding).

A	B	C	D	E	F	G	H	I	J	K
1	1	GLU	CG	2	LYS	N	0.471	2.629	INTRA	BF
2	1	GLU	OE1	34	ARG	NH1	0.174	2.376	INTRA	BF
3	1	GLU	CB	2	LYS	N	0.069	2.631	INTRA	BF
4	5	ARG	NH1	56	ASP	CG	0.392	2.708	INTRA	BF
5	5	ARG	NH2	53	HIS	O	0.3	2.25	INTRA	BF
6	5	ARG	NH1	56	ASP	OD1	0.221	2.329	INTRA	BF
7	5	ARG	NH1	56	ASP	OD2	0.154	2.396	INTRA	BF
8	6	TRP	N	36	ILE	O	0.005	2.545	INTRA	HB
9	11	ASP	OD1	39	LYS	NZ	0.005	2.545	INTRA	HB
10	14	ALA	O	15	SER	C	0.007	2.793	INTRA	
11	16	LYS	O	17	CYS	C	0.016	2.784	INTRA	
12	17	CYS	C	37	CYS	SG	0.144	3.256	INTRA	
13	24	MET	O	25	LYS	C	0.086	2.714	INTRA	BF
14	29	PRO	O	31	ASP	N	0.226	2.474	INTRA	BF
15	29	PRO	C	31	ASP	N	0.047	2.853	INTRA	BF
16	31	ASP	O	32	GLY	C	0.508	2.292	INTRA	BF
17	31	ASP	O	32	GLY	O	0.399	2.001	INTRA	BF
18	31	ASP	C	32	GLY	O	0.39	2.41	INTRA	BF
19	31	ASP	C	32	GLY	C	0.225	2.575	INTRA	BF
20	35	ILE	CG2	36	ILE	N	0.078	3.022	INTRA	BF
21	43	TYR	CG	44	LEU	N	0.174	2.926	INTRA	
22	44	LEU	O	45	ASP	C	0.019	2.781	INTRA	BF
23	49	ALA	O	54	GLU	N	0.02	2.53	INTRA	BF
24	52	ALA	O	53	HIS	CB	0.049	2.751	INTRA	BF
25	56	ASP	OD1	252	ARG	NH2	0.102	2.448	INTRA	BF
26	61	ASP	CG	62	ALA	N	0.006	3.094	INTRA	
27	71	THR	CG2	72	PRO	CA	0.112	3.088	INTRA	BF
28	72	PRO	C	74	ASN	ND2	0.238	2.862	INTRA	BF
29	73	ASN	O	74	ASN	C	0.103	2.697	INTRA	BF
30	79	VAL	CG1	80	ALA	N	0.021	3.079	INTRA	
31	103	SER	N	104	ASN	N	0.039	2.561	INTRA	B3
32	106	GLN	OE1	229	THR	C	0.011	2.789	INTRA	
33	110	LEU	O	111	GLN	C	0.114	2.686	INTRA	
34	115	SER	OG	155	SER	OG	0.031	2.219	INTRA	HB
35	116	CYS	SG	192	CYS	C	0.043	3.357	INTRA	
36	128	ILE	CB	129	PRO	CD	0.131	3.069	INTRA	
37	128	ILE	O	129	PRO	C	0.026	2.774	INTRA	
38	135	CYS	O	141	ARG	NH2	0.269	2.431	INTRA	BF
39	137	LEU	CB	138	PRO	CD	0.018	3.182	INTRA	BF
40	137	LEU	O	138	PRO	C	0.009	2.791	INTRA	BF
41	138	PRO	O	139	GLU	C	0.046	2.754	INTRA	BF
42	142	LYS	CA	143	PRO	CA	0.016	2.784	INTRA	BF
43	143	PRO	O	144	LEU	C	0.06	2.74	INTRA	BF
44	157	VAL	O	158	PRO	C	0.108	2.692	INTRA	
45	158	PRO	C	159	CYS	SG	0.112	3.288	INTRA	
46	164	ASP	C	165	PHE	CG	0.001	3.199	INTRA	BF
47	172	CYS	O	173	PRO	C	0.009	2.791	INTRA	
48	175	CYS	O	176	GLY	C	0.055	2.745	INTRA	
49	178	SER	C	180	VAL	N	0.006	2.894	INTRA	
50	183	TYR	C	188	GLY	CA	0.038	3.162	INTRA	
51	184	PHE	CG	185	GLY	N	0.083	3.017	INTRA	
52	186	TYR	OH	247	HIS	NE2	0.076	2.474	INTRA	HB
53	188	GLY	O	189	ALA	C	0.002	2.798	INTRA	
54	202	PHE	C	203	VAL	CG1	0.001	3.199	INTRA	
55	204	LYS	NZ	294	LYS	NZ	0.486	2.514	INTRA	
56	204	LYS	NZ	294	LYS	CE	0.017	3.083	INTRA	
57	206	GLU	O	207	THR	C	0.114	2.686	INTRA	
58	212	LEU	CA	213	PRO	CD	0.086	2.714	INTRA	B3
59	252	ARG	NH1	255	ASP	O	0.375	2.175	INTRA	BF
60	252	ARG	NH1	255	ASP	C	0.133	2.967	INTRA	BF
61	254	VAL	O	255	ASP	CB	0.019	2.781	INTRA	BF
62	260	LEU	O	261	ILE	C	0.002	2.798	INTRA	BF
63	263	GLU	O	267	GLN	CG	0.065	2.735	INTRA	BF
64	267	GLN	O	268	ALA	C	0.031	2.769	INTRA	
65	275	ASP	CG	276	LYS	N	0.104	2.996	INTRA	BF
66	279	ASP	N	280	PHE	N	0.02	2.58	INTRA	BF
67	284	SER	O	285	SER	CB	0.014	2.786	INTRA	
68	302	LYS	CD	303	VAL	N	0.051	3.049	INTRA	BF
69	304	PRO	CA	305	PRO	CD	0.137	2.663	INTRA	BF
70	314	GLY	O	318	VAL	CG2	0.01	2.79	INTRA	BF
71	316	GLU	O	317	TYR	C	0.021	2.779	INTRA	BF
72	323	ASN	O	324	LEU	C	0.021	2.779	INTRA	BF
73	330	PRO	C	331	ASP	O	0.212	2.588	INTRA	BF
74	331	ASP	O	333	LEU	N	0.321	2.379	INTRA	BF
75	331	ASP	O	332	PRO	C	0.185	2.615	INTRA	BF
76	331	ASP	CB	332	PRO	CD	0.022	3.178	INTRA	BF
77	334	GLN	O	335	ASP	CB	0.08	2.72	INTRA	BF
78	336	GLU	N	337	CYS	N	0.036	2.564	INTRA	BF
79	337	CYS	SG	595	LYS	CD	0.041	3.359	INTRA	BF
80	342	TRP	N	365	GLU	O	0.096	2.454	INTRA	BF
81	345	LEU	O	346	GLY	O	0.274	2.126	INTRA	BF
82	345	LEU	O	346	GLY	C	0.17	2.63	INTRA	BF
83	345	LEU	C	346	GLY	O	0.009	2.791	INTRA	BF
84	351	LEU	O	352	LYS	C	0.04	2.76	INTRA	BF
85	352	LYS	NZ	616	MET	O	0.04	2.51	INTRA	BF
86	353	CYS	C	366	CYS	SG	0.233	3.167	INTRA	BF
87	353	CYS	O	354	ASP	C	0.099	2.701	INTRA	BF
88	354	ASP	O	355	GLU	C	0.008	2.792	INTRA	BF
89	355	GLU	CG	614	PHE	CD1	0.026	3.174	INTRA	BF
90	357	SER	O	358	VAL	C	0.305	2.495	INTRA	BF
91	357	SER	C	358	VAL	C	0.029	2.771	INTRA	BF
92	370	GLU	C	371	THR	CG2	0.037	3.163	INTRA	BF
93	371	THR	CB	372	PRO	CD	0.024	3.176	INTRA	BF
94	373	GLU	OE2	663	SER	OG	0.187	2.063	INTRA	BF
95	373	GLU	O	374	ASP	C	0.009	2.791	INTRA	BF
96	375	CYS	CB	387	MET	SD	0.043	3.357	INTRA	BF
97	375	CYS	CB	387	MET	CE	0.004	3.196	INTRA	BF
98	377	ALA	CA	380	MET	CE	0.169	3.031	INTRA	BF
99	390	ASP	CG	391	GLY	N	0.058	3.042	INTRA	
100	390	ASP	O	391	GLY	C	0 014	2.786	INTRA	
101	395	TYR	O	396	ILE	C	0.022	2.778	INTRA	BF
102	396	ILE	O	397	ALA	C	0.01	2.79	INTRA	BF
103	399	GLN	N	400	CYS	N	0.019	2.581	INTRA	BF
104	405	VAL	CG1	406	LEU	N	0.416	2.684	INTRA	BF
105	415	CYS	N	632	CYS	SG	0.209	3.091	INTRA	BF
106	416	LYS	O	637	ARG	NH2	0.322	2.228	INTRA	BF
107	416	LYS	N	417	LYS	N	0.018	2.582	INTRA	BF
108	431	LYS	O	432	SER	C	0.164	2.636	INTRA	BF
109	432	SER	OG	433	ASN	N	0.135	2.565	INTRA	BF
110	437	ASN	O	438	TRP	C	0.006	2.794	INTRA	BF
111	441	LEU	O	442	GLU	C	0.171	2.629	INTRA	BF
112	445	LYS	NZ	524	ASP	OD2	0.018	2.532	INTRA	BF
113	457	TRP	O	458	ASN	C	0.107	2.693	INTRA	
114	459	ILE	N	460	PRO	CD	0.131	2.869	INTRA	
115	459	ILE	CB	460	PRO	CD	0.059	3.141	INTRA	
116	465	TYR	C	467	ARG	N	0.152	2.748	INTRA	BF
117	465	TYR	O	467	ARG	N	0.103	2.597	INTRA	BF
118	465	TYR	CE1	471	CYS	SG	0.051	3.349	INTRA	BF
119	465	TYR	CE1	660	CYS	SG	0.048	3.352	INTRA	BF
120	472	ARG	NE	475	GLU	OE1	0.197	2.353	INTRA	BF
121	472	ARG	NE	475	GLU	CD	0.077	3.023	INTRA	BF
122	482	ALA	O	483	PRO	C	0.078	2.725	INTRA	BF
123	486	GLN	O	488	ASN	N	0.198	2.502	INTRA	BF
124	486	GLN	O	487	LYS	C	0.137	2.663	INTRA	BF
125	490	SER	O	492	CYS	N	0.061	2.639	INTRA	BF
126	490	SER	C	492	CYS	N	0.024	2.876	INTRA	BF
127	492	CYS	N	493	GLU	N	0.106	2.494	INTRA	BF
128	495	CYS	C	496	VAL	CG1	0.055	3.145	INTRA	BF
129	495	CYS	CB	507	GLU	OE1	0.028	2.772	INTRA	BF
130	496	VAL	CG2	497	GLY	N	0.261	2.839	INTRA	BF
131	496	VAL	O	497	GLY	C	0.242	2.858	INTRA	BF
132	496	VAL	C	497	GLY	O	0.138	2.662	INTRA	BF
133	496	VAL	O	497	GLY	O	0.079	2.321	INTRA	BF
134	496	VAL	C	497	GLY	C	0.004	2.796	INTRA	BF
135	497	GLY	O	498	PRO	C	0.319	2.481	INTRA	BF
136	497	GLY	CA	508	GLY	N	0.051	3.049	INTRA	BF
137	497	GLY	C	498	PRO	C	0.05	2.75	INTRA	BF
138	500	VAL	O	501	CYS	CB	0.137	2.663	INTRA	BF
139	505	ASN	ND2	624	LEU	CD2	0.088	3.012	INTRA	BF
140	506	ARG	N	507	GLU	N	0.007	2.593	INTRA	BF
141	517	ARG	O	518	CYS	C	0.225	2.575	INTRA	BF
142	521	GLU	C	522	LYS	CG	0.098	3.102	INTRA	BF
143	543	GLU	O	544	PRO	C	0.091	2.709	INTRA	BF
144	550	LYS	C	552	GLU	N	0.107	2.793	INTRA	BF
145	550	LYS	O	551	GLU	C	0.061	2.739	INTRA	BF
146	550	LYS	O	552	GLU	N	0.02	2.68	INTRA	BF
147	567	SER	N	568	GLU	N	0.012	2.588	INTRA	B3
148	568	GLU	O	569	ALA	C	0.224	2.576	INTRA	
149	569	ALA	O	570	HIS	C	0.149	2.651	INTRA	
150	569	ALA	O	572	CYS	N	0.021	2.529	INTRA	HB
151	579	ASN	O	580	HIS	C	0.002	2.798	INTRA	
152	585	ARG	O	586	LYS	C	0.067	2.733	INTRA	BF
153	586	LYS	O	587	ASP	C	0.359	2.441	INTRA	BF
154	586	LYS	C	587	ASP	C	0.059	2.741	INTRA	BF
155	588	LYS	O	589	ALA	C	0.066	2.734	INTRA	BF
156	590	ALA	O	591	CYS	C	0.37	2.43	INTRA	BF
157	590	ALA	C	591	CYS	C	0.11	2.69	INTRA	BF
158	592	VAL	O	593	LYS	C	0.04	2.76	INTRA	BF
159	593	LYS	O	594	GLN	C	0.13	2.67	INTRA	BF
160	593	LYS	O	596	LEU	N	0.093	2.457	INTRA	BF
161	593	LYS	NZ	635	ASP	OD2	0.056	2.494	INTRA	BF
162	596	LEU	O	597	LEU	C	0.043	2.757	INTRA	BF
163	598	ASP	O	599	LEU	C	0.025	2.775	INTRA	BF
164	604	GLY	O	605	ASN	C	0.098	2.702	INTRA	BF
165	605	ASN	OD1	606	THR	N	0.163	2.387	INTRA	BF
166	605	ASN	CG	606	THR	N	0.092	3.008	INTRA	BF
167	607	VAL	CG1	615	CYS	SG	0.149	3.251	INTRA	BF
168	613	LYS	CG	614	PHE	N	0.141	2.959	INTRA	BF
169	619	SER	O	620	LYS	C	0.062	2.738	INTRA	BF
170	621	THR	CG2	622	LYS	N	0.551	2.549	INTRA	BF
171	641	THR	O	645	TYR	N	0.027	2.523	INTRA	BF
172	647	GLY	O	648	ALA	C	0.023	2.777	INTRA	BF
173	648	ALA	O	649	ASP	C	0.035	2.765	INTRA	BF
174	651	ILE	O	652	LYS	C	0.092	2.708	INTRA	BF
175	655	SER	O	656	ASN	C	0.017	2.783	INTRA	BF
176	660	CYS	O	661	SER	C	0.129	2.671	INTRA	BF
177	660	CYS	C	661	SER	O	0.001	2.799	INTRA	BF
178	661	SER	C	662	THR	CG2	0 08	3.12	INTRA	BF
179	662	THR	N	663	SER	N	0.149	2.451	INTRA	BF
180	671	PHE	O	672	HIS	C	0.05	2.75	INTRA	BF

I.6. Observed frequencies of amino acid replacements between closely related proteins (Dayhoff *et al* 1978).

	Gly	Ala	Val	Leu	Ile	Met	Cys	Ser	Thr	Asn	Gln	Asp	Glu	Lys	Arg	His	Phe	Tyr	Trp	Pro
Gly																				
Ala	58																			
Val	10	37																		
Leu	2	10	30																	
Ile		7	66	25																
Met	1	3	8	21	6															
Cys	1	3	3		2															
Ser	45	77	4	3	2	2	12													
Thr	5	59	19	5	13	3	1	70												
Asn	16	11	1	4	4			43	17											
Gln	3	9	3	8	1	2		5	4	5										
Asp	16	15	2		1			10	6	53	8									
Glu	11	27	4	2	4	1		9	3	9	42	83								
Lys	6	6	2	4	4	9		17	20	32	15		10							
Arg	1	3	2	2	3	2	1	14	2	2	12	9		48						
His	1	2	3	4			1	3	1	23	24	4	2	2	10					
Phe	2	2	1	17	9	2		4	1	1					1	2				
Tyr		2	2	2	1		3	2	2	4			1	1		4	26			
Trp				1				2							3		1	1		
Pro	5	35	5	4	1		1	27	7	3	9	1	4	4	7	5	1			

I.7. Peptide Library of TfR1

Whole Sequence (760 Amino acids)

MMDQARSAFSNLFGGEPLSYTRFSLARQVDGDNSHVEMKLAVDEEENADNNTKANVTKPKRCSGSICY
GTIAVIVFFLIGFMIGYLGYCKGVEPKTECERLAGTESPVREEPGEDFPAARRLYWDDLKRKLSEKLD
STDFTSTIKLLNENSYVPREAGSQKDENLALYVENQFREFKLSKVWRDQHFVKIQVKDSAQNSVIIVD
KNGRLVYLVENPGGYVAYSKAATVTGKLVHANFGTKKDFEDLYTPVNGSIVIVRAGKITFAEKVANAE
SLNAIGVLIYMDQTKFPIVNAELSFFGHAHLGTGDPYTPGFPSFNHTQFPPSRSSGLPNIPVQTISRA
AAEKLFGNMEGDCPSDWKTDSTCRMVTSESKNVKLTVSNVLKEIKILNIFGVIKGFVEPDHYVVVGAQ
RDAWGPGAAKSGVGTALLLKLAQMFSDMVLKDGFQPSRSIIFASWSAGDFGSVGATEWLEGYLSSLHL
KAFTYINLDKAVLGTSNFKVSASPLLYTLIEKTMQNVKHPVTGQFLYQDSNWASKVEKLTLDNAAFPF
LAYSGIPAVSFCFCEDTDYPYLGTTMDTYKELIERIPELNKVARAAAEVAGQFVIKLTHDVELNLDYE
RYNSQLLSFVRDLNQYRADIKEMGLSLQWLYSARGDFFRATSRLTTDFGNAEKTDRFVMKKLNDRVMR
VEYHFLSPYVSPKESPFRHVFWGSGSHTLPALLENLKLRKQNNGAFNETLFRNQLALATWTIQGAANA
LSGDVWDIDNEF

Peptides

1	...MMDQARSAFSNLF	12	LIGFMIGYLGYCK	23	REAGSQKDENLAL
2	AFSNLFGGEPLSY	13	YLGYCKGVEPKTE	24	DENLALYVENQFR
3	GEPLSYTRFSLAR	14	VEPKTECERLAGT	25	VENQFREFKLSKV
4	RFSLARQVDGDNS	15	ERLAGTESPVREE	26	FKLSKVWRDQHFV
5	VDGDNSHVEMKLA	16	SPVREEPGEDFPA	27	RDQHFVKIQVKDS
6	VEMKLAVDEEENA	17	GEDFPAARRLYWD	28	IQVKDSAQNSVII
7	DEEENADNNTKAN	18	RRLYWDDLKRKLS	29	QNSVIIVDKNGRL
8	NNTKANVTKPKRC	19	LKRKLSEKLDSTD	30	DKNGRLVYLVENP
9	TKPKRCSGSICYG	20	KLDSTDFTSTIKL	31	YLVENPGGYVAYS
10	GSICYGTIAVIVF	21	TSTIKLLNENSYV	32	..GYVAYSKAATVTG
11	IAVIVFFLIGFMI	22	NENSYVPREAGSQ	33	..AATVTGKLVHANF

34	LVHANFGTKKDFE	59	AQRDAWGPGAAKS	84	ELNKVARAAAEVA
35	TKKDFEDLYTPVN	60	PGAAKSGVGTALL	85	AAAEVAGQFVIKL
36	LYTPVNGSIVIVR	61	VGTALLLKLAQMF	86	QFVIKLTHDVELN
37	SIVIVRAGKITFA	62	KLAQMFSDMVLKD	87	HDVELNLDYERYN
38	GKITFAEKVANAE	63	DMVLKDGFQPSRS	88	DYERYNSQLLSFV
39	KVANAESLNAIGV	64	FQPSRSIIFASWS	89	QLLSFVRDLNQYR
40	LNAIGVLIYMDQT	65	IFASWSAGDFGSV	90	DLNQYRADIKEMG
41	IYMDQTKFPIVNA	66	GDFGSVGATEWLE	91	DIKEMGLSLQWLY
42	FPIVNAELSFFGH	67	ATEWLEGYLSSLH	92	SLQWLYSARGDFF
43	LSFFGHAHLGTGD	68	YLSSLHLKAFTYI	93	ARGDFFRATSRLT
44	HLGTGDPYTPGFP	69	KAFTYINLDKAVL	94	ATSRLTTDFGNAE
45	YTPGFPSFNHTQF	70	LDKAVLGTSNFKV	95	DFGNAEKTDRFVM
46	FNHTQFPPSRSSG	71	TSNFKVSASPLLY	96	TDRFVMKKLNDRV
47	PSRSSGLPNIPVQ	72	ASPLLYTLIEKTM	97	KLNDRVMRVEYHF
48	PNIPVQTISRAAA	73	LIEKTMQNVKHPV	98	RVEYHFLSPYVSP
49	ISRAAAEKLFGNM	74	NVKHPVTGQFLYQ	99	SPYVSPKESPFRH
50	KLFGNMEGDCPSD	75	GQFLYQDSNWASK	100	ESPFRHVFWGSGS
51	GDCPSDWKTDSTC	76	SNWASKVEKLTLD	101	FWGSGSHTLPALL
52	KTDSTCRMVTSES	77	EKLTLDNAAFPFL	102	TLPALLENLKLRK
53	MVTSESKNVKLTV	78	AAFPFLAYSGIPA	103	NLKLRKQNNGAFN
54	NVKLTVSNVLKEI	79	YSGIPAVSFCFCE	104	NNGAFNETLFRNQ
55	NVLKEIKILNIFG	80	SFCFCEDTDYPYL	105	TLFRNQLALATWT
56	ILNIFGVIKGFVE	81	TDYPYLGTTMDTY	106	ALATWTIQGAANA
57	IKGFVEPDHYVVV	82	TTMDTYKELIERI	107	QGAANALSGDVWD
58	DHYVVVGAQRDAW	83	ELIERIPELNKVA	108	.SGDVWDIDNEF

Appendix II. Molecular Biology

II.1. Bacterial Strains

BL21 (DE3): *hsdS, gal, (λcIts857,ind1, Sam7, nin5,lacUV5-T7 gene1)*

A strain employed for high-level expresson of genes cloned into expression vectors containing bacteriophage T7 promoter. bacteriophage T7 RNA polymerase is carried on the bacteriophage λ DE3, which is integrated into the chromosome of BL21

DH5α: 80d*lacZΔM15, recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, deoR, Δ(lac-ZYA-argF) U169, phoA*

DH5α is a well-known, versatile strain that can be used in many everyday cloning applications. In addition to supporting blue/white screening, *rec*A1 and *end*A1 mutations in DH5α is thought to increase insert stability and improve the quality of plasmid DNA prepared from minipreps.

JM109: *recA1, supE44, endA1, hsdR17, gyrA96, relA1, thi, Δ(lac-proAB),* F'*[traD36, proAB*$^{+}$*, lacI*^{q}*, lacZΔM15]*

A recombination-deficient strain that will support growth of vectors carrying a number of mutations and will modify, but not restrict, transfected DNA (Yanisch-peron et al. 1985).

Origami™

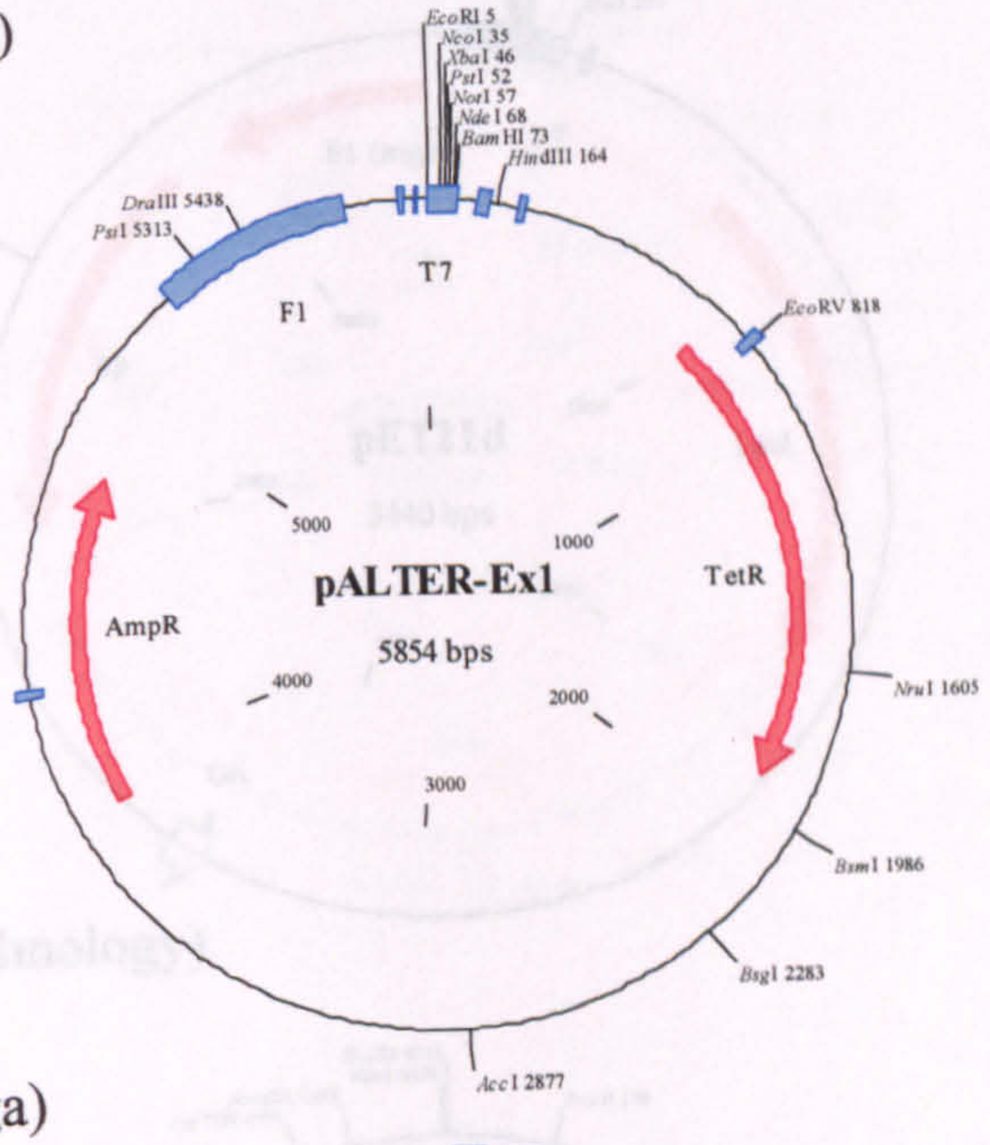
host strains are K-12 derivatives that have mutations in both the thioredoxinreductase (*trxB*) and glutathione reductase (*gor*) genes, which greatly enhances disulfide bondformation in the cytoplasm (13). Studies have shown that expression in Origami(DE3) yielded 10-fold more active protein than in a *trxB* single mutant host even though overall expression levels were similar(Prinz *et al.*, 1997). Origami hosts are not compatible with kanamycin or tetracycline resistantplasmids. Origami hosts are ideal for use with pET-32 vectors, since the thioredoxin fusion tag further enhances the formation of disulfide bonds in the cytoplasm. Origami™ (DE3) strains are the same as Origami but are lysogenic for a λ prophage that contains an IPTG inducible T7 RNA polymerase.

XL-1 Blue: *recA1, endA1, gyrA96, thi-1, hsdR17(*r_k*-,*m_k*+), supE44, relA1, lac,* {*f*'*, proAB, lacI*^{q}*ZΔM15::Tn10(tet*^{r}*)]*

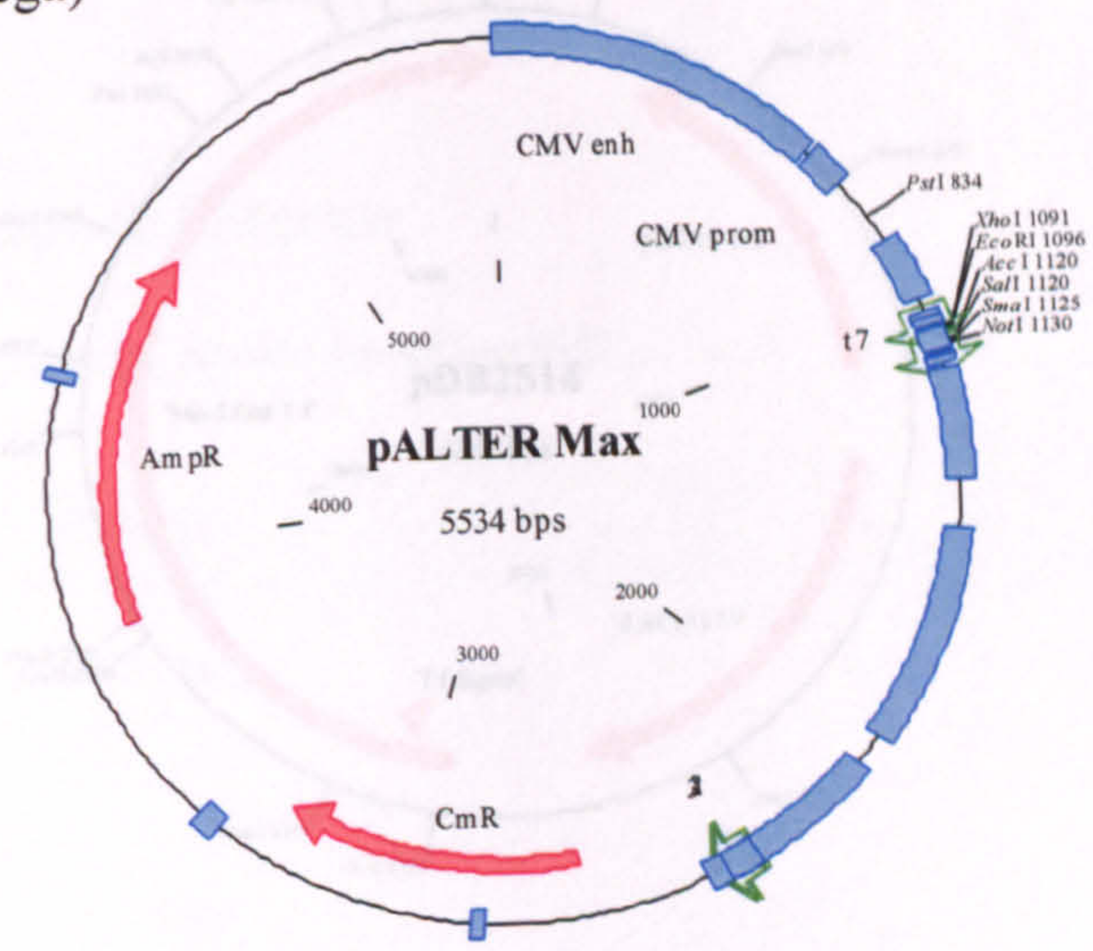
Strain of choice for preparation of high-quality plasmid DNA, or use in difficult transformations. Allows blue-white color screening and single-strand rescue of phagemid DNA. Contains an tetracycline-resistant F' episome, eliminating time-consuming selection on minimal media plates.

II.2. Vectors

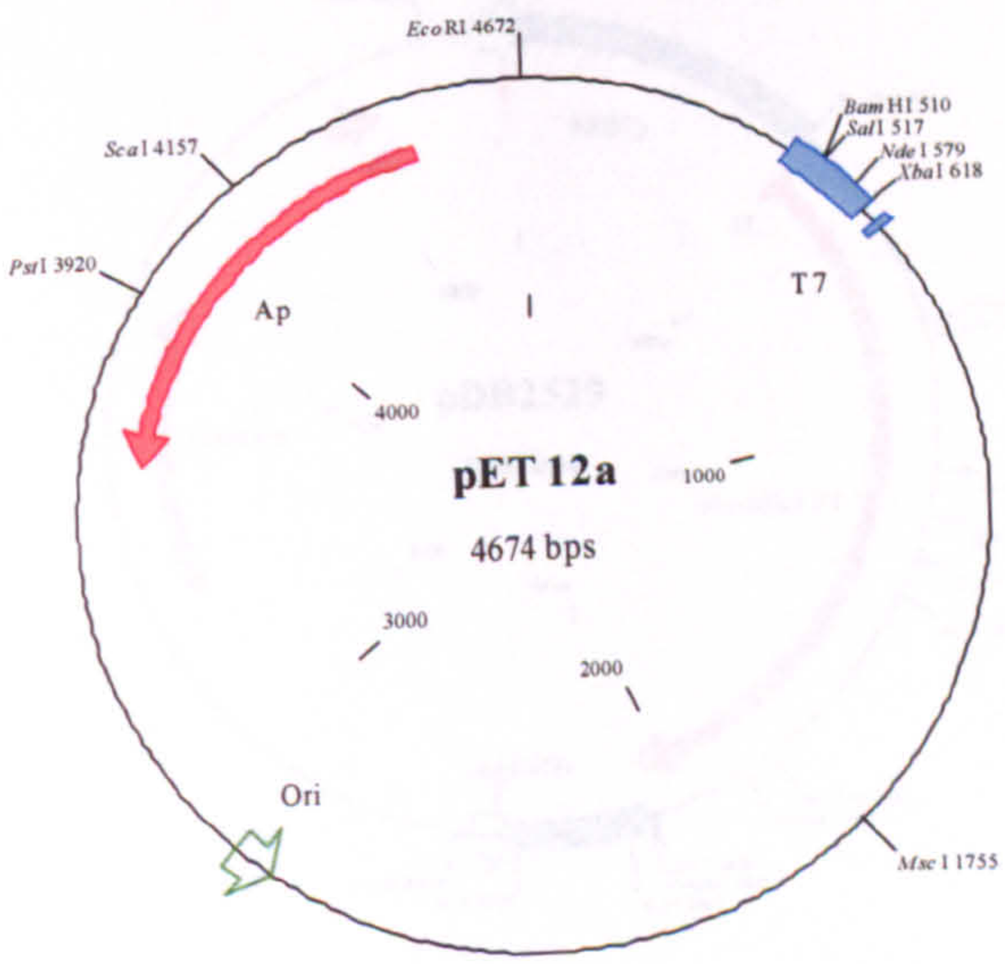
pALTER-Ex1 (Promega)



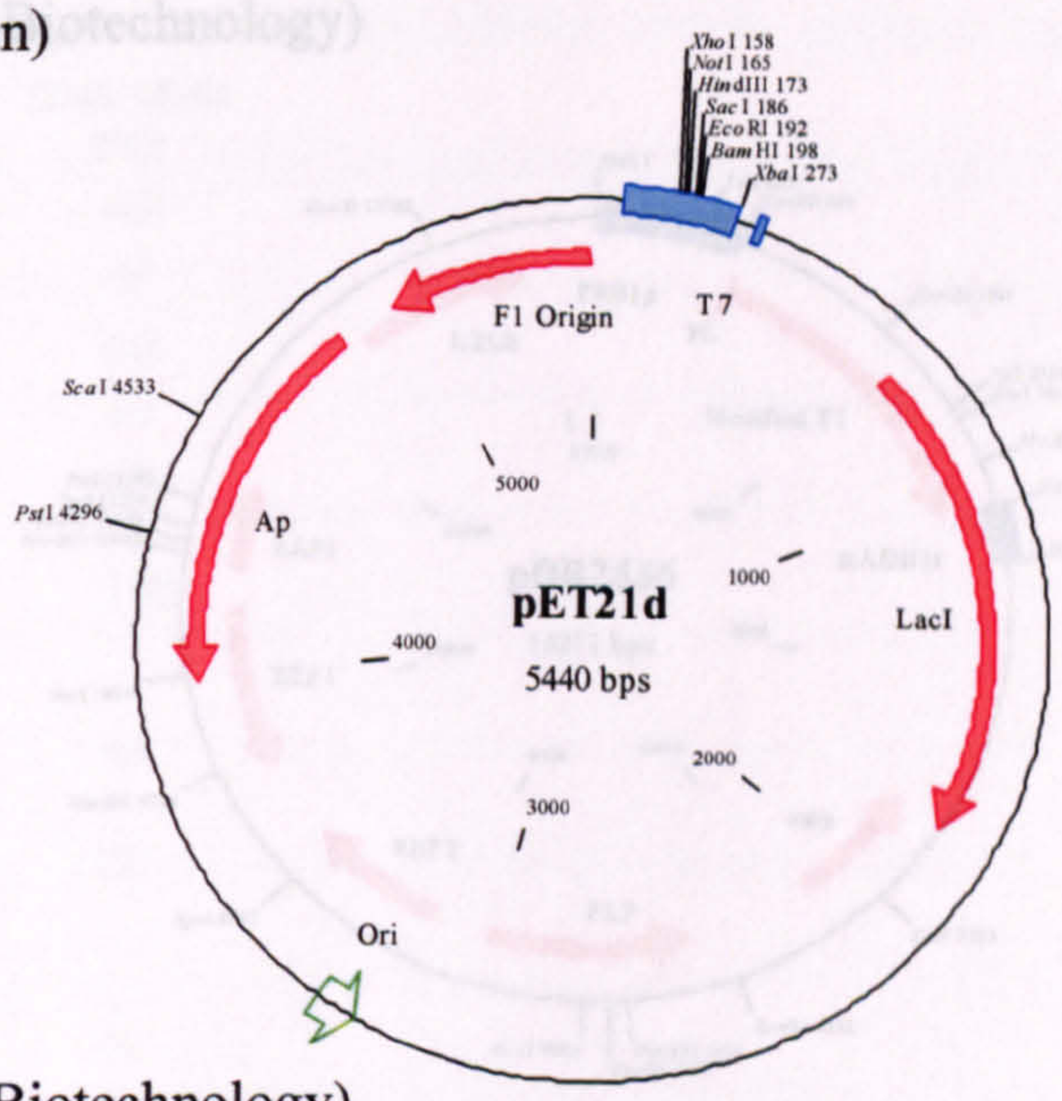
pALTER-max (Promega)



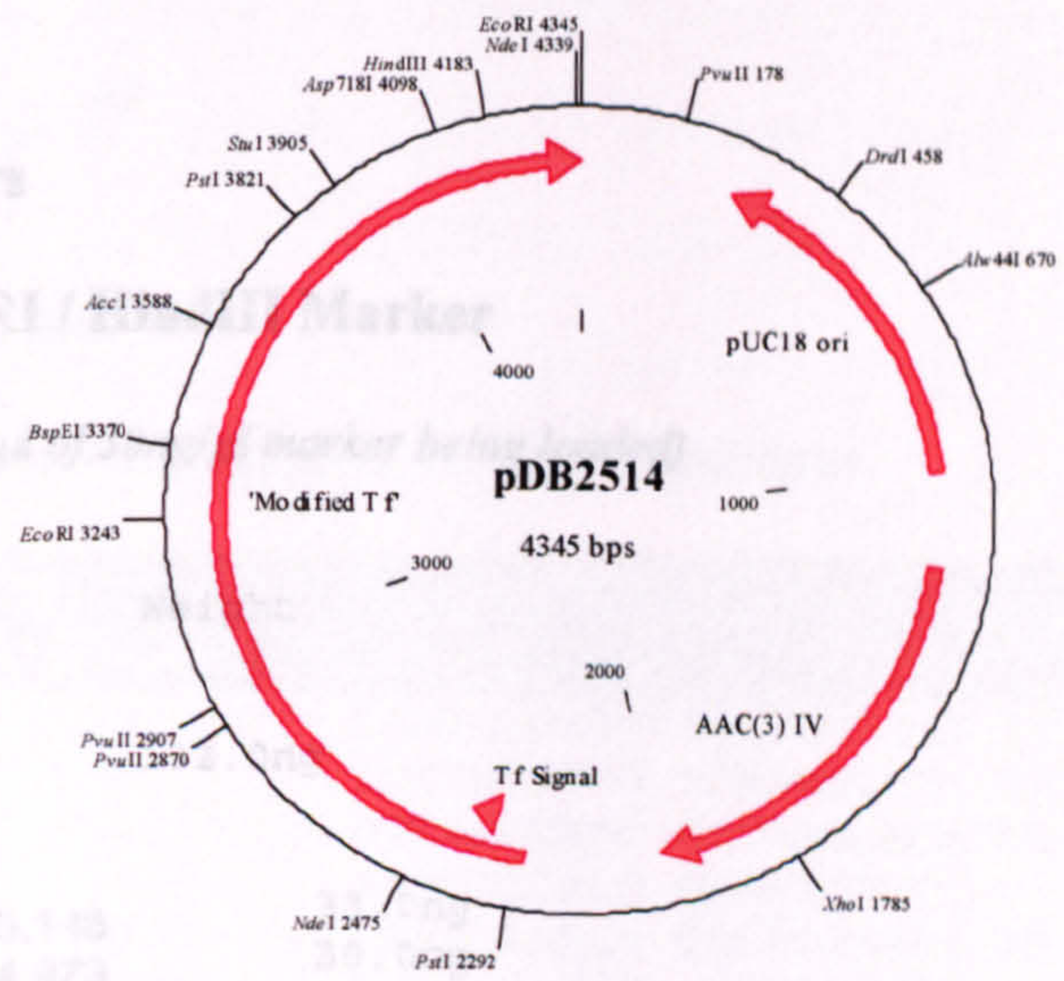
pET12a (Novagen)



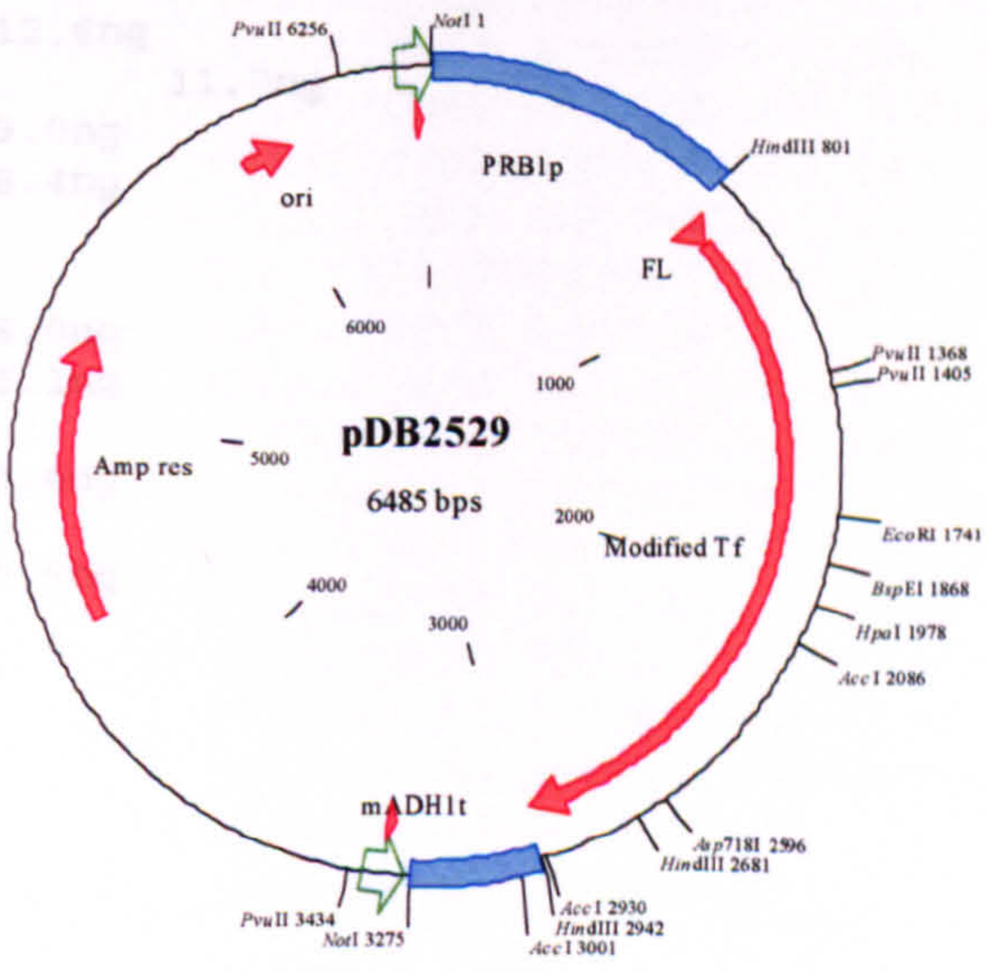
pET21d (Novagen)



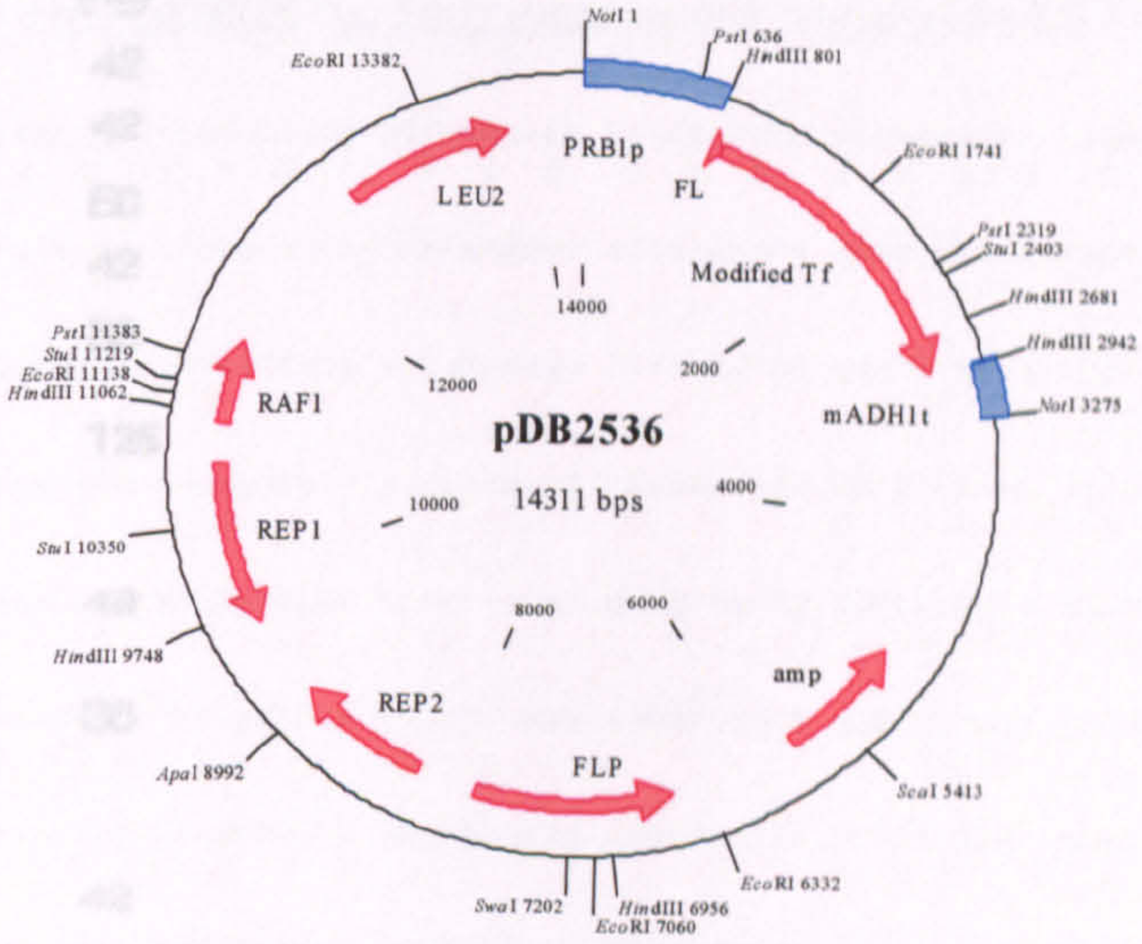
pDB2514 (Delta Biotechnology)



pDB2529 (Delta Biotechnology)



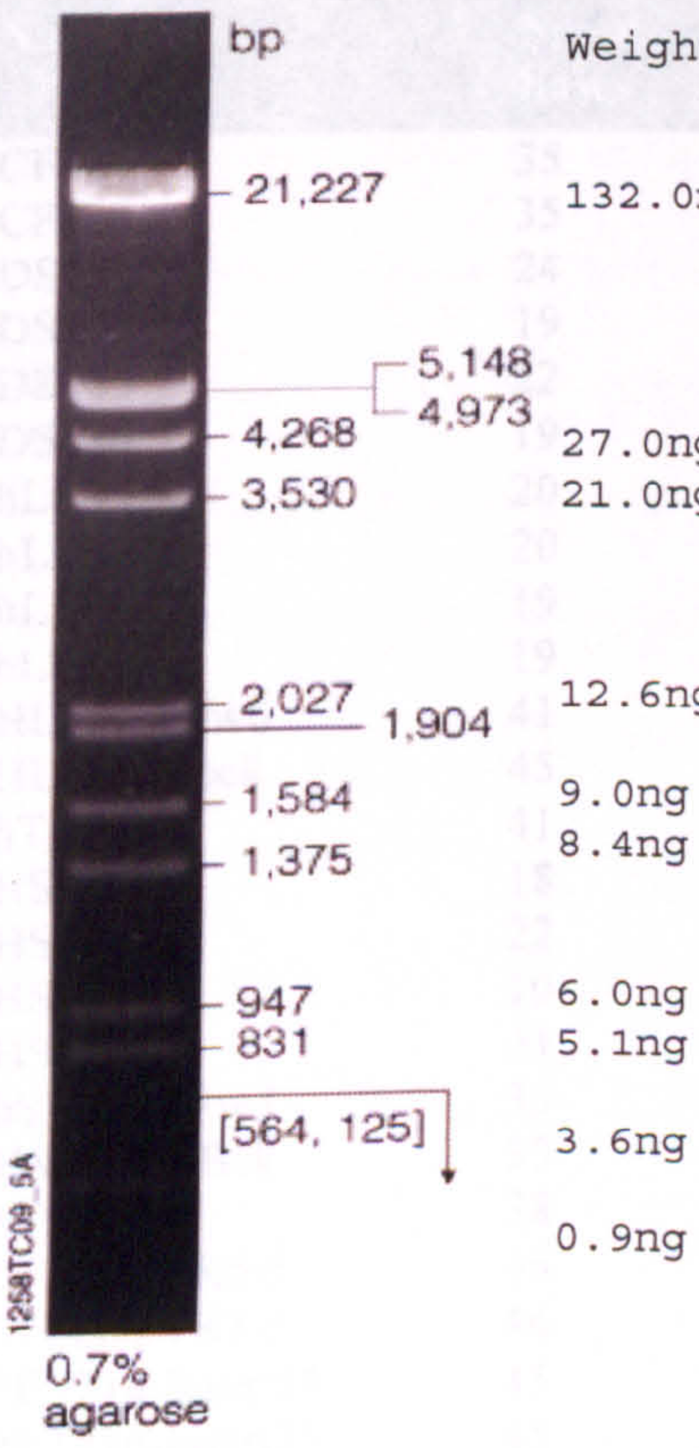
pDB2536 (Delta Biotechnology)



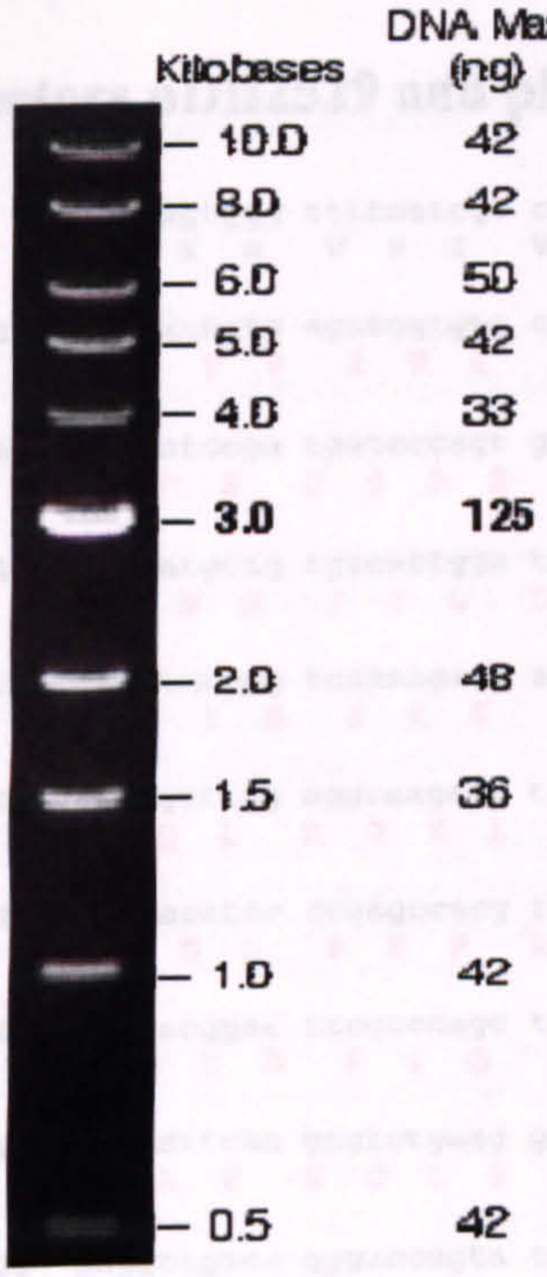
II.3. DNA ladders

i. Lambda / EcoRI / HindIII Marker

(weights based on 10µl of 30ng/µl marker being loaded)



ii. 1Kb DNA ladder



II.4. Oligonucleotides

All oligonucleotides were supplied by Sigma-Genosys.

Name	No. of Bases	Sequence (5' – 3')
CF124	35	CCATGAGCTT GGATGGAAGA TTTGTCTACA TAGCG
CF125	35	CGCTATGTAG ACAAATCTTC CATCCAAGCT CATGG
DS181	24	CTCAACCAGG CCCAGGAACA TTTT
DS182	19	AGAGACCACC GAAGACTGC
DS185	22	GACAGAAACA CATATGAAAA AT
DS186	19	GGAGATGTGG CCTTTGTGA
hLFR1	20	AGCACAGAGA TTAGATCTCG
hLFF1	20	ACAGGGGAAA ACAAATGTGC
hLFF2	19	ATCATGCCGT GGTGTCTCG
hLFR2	19	TCTTTATGCA GCTGACAGG
HLf PCR fwd	41	GGAATTCAGT ACTGGATCCA TGGGTCGACG
HLf PCR bck	45	TCCGCTCGAG CAATTGAAGC TTTACTTCCT
hTrfFwd	41	GGAATTCTCG AGGCCGCCAC CATGAGGCTC
HSTFR1	18	AGCAGCCACA CCCTGGAC
HSTFF1	22	TTACTTTACT GTGACTTACC TG
HSTFF2	19	ATGGATGCCA AGATGTACC
HSTFF3	21	AACAACAAAG AGGGATACTA C
hSTfp21dFwd	33	CGCGGATCCA TGGTCCCTGA TAAAACTGTG AGA
HSTfp21dBck	33	CTGAAGATCT TCTTTTGACC CATAGAACTC TGC
hTrfBck	38	TTTTCCTTTT GCGGCCGCTT AAGGTCTACG GAAAGTGC
LFPompTK5-3	46	CTAGAAATAA TTTTGTTTAA CTTTAAGAAG
LFPompTK3-5	46	TCGACCCATA TGTATATCTC CTTCTTAAAG
PET21d-insert53	45	CTAGAAATAA TTTTGTTTAA CTTTAAGAAG
PET21d-insert35	45	TCGACCATGG ATCCATCTCC TTCTTAAAGT
TfG394R-CKF	20	GCTGTTTAAG GACTCTGCCC
TfG394R-CKR	20	AGACAGCAAA ATACCCTGCC

II.5. The hSTf (N413Q, N611Q) cDNA insert present in the yeast expression vectors pDB2529 and pDB2536. Sequence in red highlights hSTf.

```
1    atgaagtggg ttttcatcgt ctccattttg ttcttgttct cctctgctta ctctagatct ttggataaga gagtccctga
       M  K  W     V  F  I     V  S  I  L     F  L  F     S  S  A     Y  S  R  S     L  D  K     R  V  P

81   taaaactgtg agatggtgtg cagtgtcgga gcatgaggcc actaagtgcc agagtttccg cgaccatatg aaaagcgtca
     D  K  T  V     R  W  C     A  V  S     E  H  E  A     T  K  C     Q  S  F     R  D  H  M     K  S  V

161  ttccatccga tggtcccagt gttgcttgtg tgaagaaagc ctcctacctt gattgcatca gggccattgc ggcaaacgaa
       I  P  S     D  G  P  S     V  A  C     V  K  K     A  S  Y  L     D  C  I     R  A  I     A  A  N  E

241  gcggatgctg tgacactgga tgcaggtttg gtgtatgatg cttacctggc tcccaataac ctgaagcctg tggtggcaga
       A  D  A     V  T  L     D  A  G  L     V  Y  D     A  Y  L     A  P  N  N     L  K  P     V  V  A

321  gttctatggg tcaaaagagg atccacagac tttctattat gctgttgctg tggtgaagaa ggatagtggc ttccagatga
       E  F  Y  G     S  K  E     D  P  Q     T  F  Y  Y     A  V  A     V  V  K     K  D  S  G     F  Q  M

401  accagcttcg aggcaagaag tcctgccaca cgggtctagg caggtccgct gggtggaaca tccccatagg cttactttac
       N  Q  L     R  G  K  K     S  C  H     T  G  L     G  R  S  A     G  W  N     I  P  I     G  L  L  Y

481  tgtgacttac ctgagccacg taaacctctt gagaaagcag tggccaattt cttctcgggc agctgtgccc cttgtgcgga
       C  D  L     P  E  P     R  K  P  L     E  K  A     V  A  N     F  F  S  G     S  C  A     P  C  A

561  tgggacggac ttcccccagc tgtgtcaact gtgtccaggg tgtggctgct ccacccttaa ccaatacttc ggctactcgg
       D  G  T  D     F  P  Q     L  C  Q     L  C  P  G     C  G  C     S  T  L     N  Q  Y  F     G  Y  S

641  gagccttcaa gtgtctgaag gatggtgctg gggatgtggc ctttgtcaag cactcgacta tatttgagaa cttggcaaac
       G  A  F     K  C  L  K     D  G  A     G  D  V     A  F  V  K     H  S  T  I  F  E     N  L  A  N

721  aaggctgaca gggaccagta tgagctgctt tgcctggaca acacccggaa gccggtagat gaatacaagg actgccactt
       K  A  D     R  D  Q     Y  E  L  L     C  L  D     N  T  R     K  P  V  D     E  Y  K     D  C  H

801  ggcccaggtc ccttctcata ccgtcgtggc ccgaagtatg ggcggcaagg aggacttgat ctgggagctt ctcaaccagg
       L  A  Q  V     P  S  H     T  V  V     A  R  S  M     G  G  K     E  D  L     I  W  E  L     L  N  Q

881  cccaggaaca ttttggcaaa gacaaatcaa aagaattcca actattcagc tctcctcatg ggaaggacct gctgtttaag
       A  Q  E     H  F  G  K     D  K  S     K  E  F     Q  L  F  S     S  P  H     G  K  D     L  L  F  K

961  gactctgccc acgggttttt aaaagtcccc cccaggatgg atgccaagat gtacctgggc tatgagtatg tcactgccat
       D  S  A     H  G  F     L  K  V  P     P  R  M     D  A  K     M  Y  L  G     Y  E  Y     V  T  A

1041 ccggaatcta cgggaaggca catgcccaga agccccaaca gatgaatgca agcctgtgaa gtggtgtgcg ctgagccacc
       I  R  N  L     R  E  G     T  C  P     E  A  P  T     D  E  C     K  P  V     K  W  C  A     L  S  H

1121 acgagaggct caagtgtgat gagtggagtg ttaacagtgt agggaaaata gagtgtgtat cagcagagac caccgaagac
       H  E  R     L  K  C  D     E  W  S     V  N  S     V  G  K  I     E  C  V     S  A  E     T  T  E  D

1201 tgcatcgcca agatcatgaa tggagaagct gatgccatga gcttggatgg agggtttgtc tacatagcgg gcaagtgtgg
       C  I  A     K  I  M     N  G  E  A     D  A  M     S  L  D     G  G  F  V     Y  I  A     G  K  C

1281 tctggtgcct gtcttggcag aaaactacca aaagagcgat aattgtgagg atacaccaga ggcagggtat tttgctgtag
       G  L  V  P     V  L  A     E  N  Y     Q  K  S  D     N  C  E     D  T  P     E  A  G  Y     F  A  V

1361 cagtggtgaa gaaatcagct tctgacctca cctgggacaa tctgaaaggc aagaagtcct gccatacggc agttggcaga
       A  V  V     K  K  S  A     S  D  L     T  W  D     N  L  K  G     K  K  S     C  H  T     A  V  G  R

1441 accgctggct ggaacatccc catgggcctg ctctacaata agatcaacca ctgcagattt gatgaatttt tcagtgaagg
       T  A  G     W  N  I     P  M  G  L     L  Y  N     K  I  N     H  C  R  F     D  E  F     F  S  E

1521 ttgtgcccct gggtctaaga aagactccag tctctgtaag ctgtgtatgg gctcaggcct aaacctctgt gaacccaaca
       G  C  A  P     G  S  K     K  D  S     S  L  C  K     L  C  M     G  S  G     L  N  L  C     E  P  N

1601 acaaagaggg atactacggc tacacaggcg ctttcaggtg tctggttgag aagggagatg tggcctttgt gaaacaccag
       N  K  E     G  Y  Y  G     Y  T  G     A  F  R     C  L  V  E     K  G  D     V  A  F     V  K  H  Q

1681 actgtcccac agaacactgg gggaaaaaac cctgatccat gggctaagaa tctgaatgaa aaagactatg agttgctgtg
       T  V  P     Q  N  T     G  G  K  N     P  D  P     W  A  K     N  L  N  E     K  D  Y     E  L  L

1761 ccttgatggt accaggaaac ctgtggagga gtatgcgaac tgccacctgg ccagagcccc gaatcacgct gtggtcacac
       C  L  D  G     T  R  K     P  V  E     E  Y  A  N     C  H  L     A  R  A     P  N  H  A     V  V  T

1841 ggaaagataa ggaagcttgc gtccacaaga tattacgtca acagcagcac ctatttggaa gccaagtaac tgactgctcg
       R  K  D     K  E  A  C     V  H  K     I  L  R     Q  Q  Q  H     L  F  G     S  Q  V     T  D  C  S

1921 ggcaactttt gtttgttccg gtcggaaacc aaggaccttc tgttcagaga tgacacagta tgtttggcca aacttcatga
       G  N  F     C  L  F     R  S  E  T     K  D  L     L  F  R     D  D  T  V     C  L  A     K  L  H

2001 cagaaacaca tatgaaaaat acttaggaga agaatatgtc aaggctgttg gtaacctgag aaaatgctcc acctcatcac
       D  R  N  T     Y  E  K     Y  L  G     E  E  Y  V     K  A  V     G  N  L     R  K  C  S     T  S  S

2081 tcctggaagc ctgcactttc cgtagacctt aataa
       L  L  E     A  C  T  F     R  R  P     -
```

Appendix III: Crystallisation

III.1. Crystal Screen 1: Hampton Research

	Salt	Buffer	pH	Precipitant
1	0 02M Calcium chloride dihydrate	0.1M Sodium acetate trihydrate	4.6	30 %v/v 2-methyl-2,4-pentanediol
2				0 4M Potassium sodium tartrate tetrahydrate
3				0 4M Ammonium dihydrogen phosphate
4		0.1M Tris hydrochloride	8.5	2M Ammonium sulfate
5	0.2M tri-Sodium citrate dihydrate	0.1M Sodium HEPES	7.5	30 %v/v 2-methyl-2,4-pentanediol
6	0 2M magnesium chloride hexahydrate	0.1M Tris hydrochloride	8 5	30 %w/v PEG 4000
7		0 1M sodium cacodylate	6.5	1.4M sodium acetate trihydrate
8	0.2M tri-sodium citrate dihydrate	0.1M sodium cacodylate	6.5	30 %v/v iso-propanol
9	0 2M ammonium acetate	0.1M tri-sodium citrate dihydrate	5.6	30 %w/v PEG 4000
10	0 2M ammonium acetate	0.1M Sodium acetate trihydrate	4 6	30 %w/v PEG 4000
11		0.1M tri-sodium citrate dihydrate	5.6	1M Ammonium dihydrogen phosphate
12	0.2M magnesium chloride hexahydrate	0.1M Sodium HEPES	7.5	30 %v/v iso-propanol
13	0 2M tri-Sodium citrate dihydrate	0.1M Tris hydrochloride	8.5	30 %v/v PEG 400
14	0 2M Calcium chloride dihydrate	0.1M Sodium HEPES	7.5	28 %v/v PEG 400
15	0 2M ammonium sulfate	0.1M sodium cacodylate	6.5	30 %w/v PEG 8000
16		0 1M Sodium HEPES	7.5	1 5M lithium sulfate monohydrate
17	0 2M lithium sulfate monohydrate	0.1M Tris hydrochloride	8 5	30 %w/v PEG 4000
18	0.2M magnesium acetate tetrahydrate	0 1M sodium cacodylate	6.5	20 %w/v PEG 8000
19	0.2M ammonium acetate	0.1M Tris hydrochloride	8.5	30 %v/v iso-propanol
20	0.2M ammonium sulfate	0.1M Sodium acetate trihydrate	4.6	25 %w/v PEG 4000
21	0.2M magnesium acetate tetrahydrate	0.1M sodium cacodylate	6.5	30 %v/v 2-methyl-2,4-pentanediol
22	0 2M sodium acetate trihydrate	0.1M Tris hydrochloride	8 5	30 %w/v PEG 4000
23	0 2M magnesium chloride hexahydrate	0.1M Sodium HEPES	7.5	30 %v/v PEG 400
24	0.2M Calcium chloride dihydrate	0.1M Sodium acetate trihydrate	4.6	20 %v/v iso-propanol
25		0.1M imidazole	6.5	1M sodium acetate trihydrate
26	0 2M ammonium acetate	0 1M tri-sodium citrate dihydrate	5.6	30 %v/v 2-methyl-2,4-pentanediol
27	0 2M tri-Sodium citrate dihydrate	0 1M Sodium HEPES	7 5	20 %v/v iso-propanol
28	0.2M sodium acetate trihydrate	0.1M sodium cacodylate	6.5	30 %w/v PEG 8000
29		0.1M Sodium HEPES	7 5	0 8M Potassium sodium tartrate tetrahydrate
30	0 2M ammonium sulfate			30 %w/v PEG 8000
31	0.2M ammonium sulfate			30 %w/v PEG 4000
32				2M Ammonium sulfate
33				4M sodium formate
34		0 1M sodium acetate trihydrate	4 6	2M sodium formate
35		0.1M Sodium HEPES	7.5	0.8M sodium dihydrogen phosphate, 0.8M potassium dihydrogen phosphate
36		0.1M Tris hydrochloride	8 5	8%w/v PEG 8000
37		0 1M Sodium acetate trihydrate	4 6	8 %w/v PEG 4000
38		0 1M Sodium HEPES	7 5	1.4M tri-sodium citrate dihydrate
39		0 1M Sodium HEPES	7.5	2 %v/v PEG 400, 2M ammonium sulfate
40		0 1M tri-sodium citrate dihydrate	5.6	20 %v/v iso-propanol, 20% (w/v) PEG 4000
41		0 1M Sodium HEPES	7 5	10 %v/v iso-propanol, 20% (w/v) PEG 4000
42	0.05M potassium dihydrogen phosphate			20 %w/v PEG 8000
43				30 %w/v PEG 1500
44				0.2M magnesium formate
45	0 2M zinc acetate dihydrate	0.1M sodium cacodylate	6.5	18 %w/v PEG 8000
46	0.2M calcium acetate hydrate	0.1M sodium cacodylate	6 5	18 %w/v PEG 8000
47		0.1M Sodium acetate trihydrate	4 6	2M Ammonium sulfate
48		0 1M Tris hydrochloride	8 5	2M Ammonium dihydrogen phosphate
49	1M lithium sulfate monohydrate			2%w/v PEG 8000
50	0.5M lithium sulfate monohydrate			15%w/v PEG 8000

III.2. Crystal Screen 2: Hampton Research

	Salt	Buffer	pH	Precipitant
1	2M sodium chloride			10%w/v PEG 6000
2	0.5M sodium chloride 0 01M magnesium chloride hexahydrate			0.01M hexadecyltrimethylammonium bromide
3				25%v/v ethylene glycol
4				35%v/v dioxane
5	2M ammonium sulfate			5%v/v iso-propanol
6			7.0	1M imidazole
7				10%w/v PEG 1000, 10%w/v PEG8000
8	1 5M sodium chloride			10%v/v ethanol
9		0.1M sodium acetate trihydrate	4 6	2M sodium chloride
10	0 2M sodium chloride	0.1M sodium acetate trihydrate	4 6	30%v/v MPD
11	0 01M cobaltous chloride hexahydrate	0.1M sodium acetate trihydrate	4.6	1M 1,6 hexanediol
12	0 1M cadmium chloride dihydrate	0.1M sodium acetate trihydrate	4 6	30%w/v PEG 400
13	0 2M ammonium sulfate	0.1M sodium acetate trihydrate	4.6	30%w/v PEG monomethyl ether 2000
14	0 2M potassium sodium tartrate tetrahydrate	0.1M tri-sodium citrate dihydrate	5.6	2M ammonium sulfate
15	0 5M ammonium sulfate	0 1M tri-sodium citrate dihydrate	5.6	1M lithium sulfate monohydrate
16	0 5M sodium chloride	0.1M tri-sodium citrate dihydrate	5.6	2%v/v ethylene imine polymer
17		0.1M tri-sodium citrate dihydrate	5.6	35%v/v tert-butanol
18	0.01M ferric chloride hexahydrate	0.1M tri-sodium citrate dihydrate	5.6	10%v/v jeffamine M-600
19		0.1M tri-sodium citrate dihydrate	5.6	2.5M 1,6 hexanediol
20		0.1M MES	6.5	1.6M magnesium sulfate heptahydrate
21	0 1M sodium dihydrogen phosphate 0.1M potassium dihydrogen phosphate	0.1M MES	6.5	2M sodium chloride
22		0.1M MES	6.5	12%w/v PEG 20,000
23	1 6M ammonium sulfate	0 1M MES	6.5	10%v/v dioxane
24	0 05M cesium chloride	0.1M MES	6.5	30%v/v jeffamine M-600
25	0.01M cobaltous chloride hexahydrate	0.1M MES	6.5	1.8M ammonium sulfate
26	0.2M ammonium sulfate	0 1M MES	6.5	30%w/v PEG monomethyl ether 5000
27	0 01M zinc sulfate heptahydrate	0.1M MES	6.5	25%v/v PEG monomethyl ether 550
28			6.5	1.6M tri-sodium citrate dihydrate
29	0 5M ammonium sulfate	0.1M HEPES	7.5	30%v/v MPD
30		0.1M HEPES	7.5	10%w/v PEG 6000, 5%v/v MPD
31		0 1M HEPES	7.5	20%v/v jeffamine M-600
32	0 1M sodium chloride	0 1M HEPES	7.5	1.6M ammonium sulfate
33		0.1M HEPES	7.5	2M ammonium formate
34	0 05M cadmium sulfate hydrate	0 1M HEPES	7.5	1M sodium acetate trihydrate
35		0 1M HEPES	7.5	70%v/v MPD
36		0.1M HEPES	7.5	4.3M sodium chloride
37		0.1M HEPES	7.5	10%w/v PEG 8000
38		0.1M HEPES	7.5	20%w/v PEG 10,000, 8%v/v ethylene glycol
39	0.2M magnesium chloride hexahydrate	0 1M tris	8 5	3.4M 1,6 hexanediol
40		0.1M tris	8.5	25%v/v tert-butanol
41	0 01M nickel (II) chloride hexahydrate	0.1M tris	8.5	1M lithium sulfate monohydrate
42	1 5M ammonium sulfate	0 1M tris	8.5	12%v/v glycerol anhydrous
43	0 2M ammonium dihydrogen phosphate	0 1M tris	8 5	50%v/v MPD
44		0.1M tris	8.5	20%v/v ethanol
45	0 01M nickel (II) chloride hexahydrate	0 1M tris	8.5	20%w/v PEG monomethyl ether 2000
46	0 1M sodium chloride	0.1M bicine	9.0	20%v/v PEG monomethyl ether 550
47		0.1M bicine	9.0	2M magnesium chloride hexahydrate
48		0.1M bicine	9.0	2%v/v dioxane, 10%w/v PEG20000

III.3. Crystallisation Screen 3: hLf-In^{3+}

	Salt	Buffer	pH	Precipitant
1		0 01M Tris Base	7.5	5% (v/v) Ethanol
2		0.01M Tris Base	7.5	10% (v/v) Ethanol
3		0.01M Tris Base	7.5	15% (v/v) Ethanol
4		0.01M Tris Base	7 5	20% (v/v) Ethanol
5		0.01M Tris Base	7.5	25% (v/v) Ethanol
6		0.01M Tris Base	7.5	30% (v/v) Ethanol
7		0.01M Tris Base	8.0	5% (v/v) Ethanol
8		0.01M Tris Base	8.0	10% (v/v) Ethanol
9		0.01M Tris Base	8.0	15% (v/v) Ethanol
10		0.01M Tris Base	8.0	20% (v/v) Ethanol
11		0.01M Tris Base	8 0	25% (v/v) Ethanol
12		0.01M Tris Base	8.0	30% (v/v) Ethanol
13		0 01M Tris Base	8.5	5% (v/v) Ethanol
14		0.01M Tris Base	8.5	10% (v/v) Ethanol
15		0.01M Tris Base	8.5	15% (v/v) Ethanol
16		0 01M Tris Base	8 5	20% (v/v) Ethanol
17		0 01M Tris Base	8.5	25% (v/v) Ethanol
18		0.01M Tris Base	8.5	30% (v/v) Ethanol
19		0.01M Tris Base	9 0	5% (v/v) Ethanol
20		0 01M Tris Base	9.0	10% (v/v) Ethanol
21		0 01M Tris Base	9.0	15% (v/v) Ethanol
22		0.01M Tris Base	9.0	20% (v/v) Ethanol
23		0 01M Tris Base	9.0	25% (v/v) Ethanol
24		0.01M Tris Base	9.0	30% (v/v) Ethanol
25		0.01M Tris Base	7.5	5% (v/v) Iso-propanol
26		0.01M Tris Base	7.5	10% (v/v) Iso-propanol
27		0 01M Tris Base	7.5	15% (v/v) Iso-propanol
28		0 01M Tris Base	7 5	20% (v/v) Iso-propanol
29		0.01M Tris Base	7.5	25% (v/v) Iso-propanol
30		0.01M Tris Base	7.5	30% (v/v) Iso-propanol
31		0 01M Tris Base	8.0	5% (v/v) Iso-propanol
32		0 01M Tris Base	8 0	10% (v/v) Iso-propanol
33		0.01M Tris Base	8.0	15% (v/v) Iso-propanol
34		0 01M Tris Base	8.0	20% (v/v) Iso-propanol
35		0.01M Tris Base	8 0	25% (v/v) Iso-propanol
36		0.01M Tris Base	8 0	30% (v/v) Iso-propanol
37		0 01M Tris Base	8 5	5% (v/v) Iso-propanol
38		0.01M Tris Base	8.5	10% (v/v) Iso-propanol
39		0.01M Tris Base	8 5	15% (v/v) Iso-propanol
40		0.01M Tris Base	8.5	20% (v/v) Iso-propanol
41		0.01M Tris Base	8.5	25% (v/v) Iso-propanol
42		0.01M Tris Base	8.5	30% (v/v) Iso-propanol
43		0.01M Tris Base	9.0	5% (v/v) Iso-propanol
44		0.01M Tris Base	9.0	10% (v/v) Iso-propanol
45		0.01M Tris Base	9.0	15% (v/v) Iso-propanol
46		0.01M Tris Base	9.0	20% (v/v) Iso-propanol
47		0.01M Tris Base	9.0	25% (v/v) Iso-propanol
48		0.01M Tris Base	9.0	30% (v/v) Iso-propanol
49		0 01M Tris Base	7.5	5% (v/v) PEG 6000
50		0 01M Tris Base	7.5	15% (v/v) PEG 6000
51		0 01M Tris Base	7.5	20% (v/v) PEG 6000
52		0 01M Tris Base	7.5	25% (v/v) PEG 6000
53		0 01M Tris Base	7.5	30% (v/v) PEG 6000
54		0.01M Tris Base	7 5	35% (v/v) PEG 6000
55		0 01M Tris Base	8 0	5% (v/v) PEG 6000
56		0 01M Tris Base	8.0	15% (v/v) PEG 6000
57		0 01M Tris Base	8.0	20% (v/v) PEG 6000
58		0 01M Tris Base	8 0	25% (v/v) PEG 6000
59		0 01M Tris Base	8.0	30% (v/v) PEG 6000
60		0 01M Tris Base	8 0	35% (v/v) PEG 6000
61		0 01M Tris Base	8.5	5% (v/v) PEG 6000
62		0 01M Tris Base	8 5	15% (v/v) PEG 6000
63		0 01M Tris Base	8.5	20% (v/v) PEG 6000
64		0.01M Tris Base	8 5	25% (v/v) PEG 6000
65		0 01M Tris Base	8.5	30% (v/v) PEG 6000
66		0.01M Tris Base	8 5	35% (v/v) PEG 6000
67		0 01M Tris Base	9 0	5% (v/v) PEG 6000
68		0 01M Tris Base	9 0	15% (v/v) PEG 6000
69		0 01M Tris Base	9.0	20% (v/v) PEG 6000
70		0 01M Tris Base	9.0	25% (v/v) PEG 6000
71		0.01M Tris Base	9 0	30% (v/v) PEG 6000
72		0.01M Tris Base	9 0	35% (v/v) PEG 6000

III.4. Crystallisation of hLf-In^{3+}: Screen 1

No Crystals	Precipitate	Crystals (3 weeks)	Crystals (4 weeks)	Crystals (5 weeks)

Ethanol

P1 10mg/ml

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24

P2 20mg/ml

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24

P3 30mg/ml

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24

P4 40mg/ml

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24

P5 10mg/ml

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24

P6 20mg/ml

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24

P7 30mg/ml

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24

P8 40mg/ml

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24

Isopropanol

P9 10mg/ml

25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48

P10 20mg/ml

25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48

P11 30mg/ml

25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48

P12 40mg/ml

25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48

P13 10mg/ml

25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48

P14 20mg/ml

25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48

P15 30mg/ml

25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48

P16 40mg/ml

25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48

PEG6000

P17	10mg/ml				
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	66
67	68	69	70	71	72

P18	20mg/ml				
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	66
67	68	69	70	71	72

P19	30mg/ml				
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	66
67	68	69	70	71	72

P20	40mg/ml				
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	66
67	68	69	70	71	72

P21	10mg/ml				
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	66
67	68	69	70	71	72

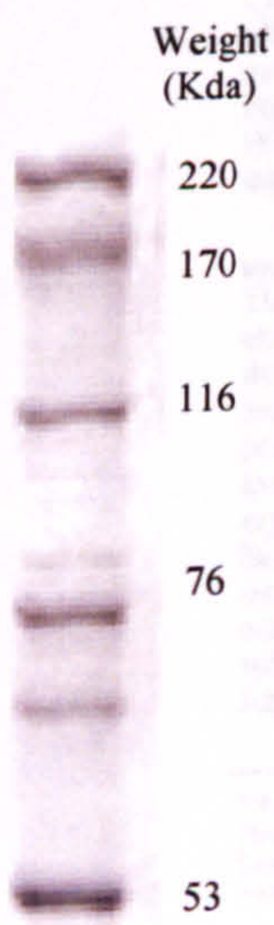
P22	20mg/ml				
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	66
67	68	69	70	71	72

P23	30mg/ml				
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	66
67	68	69	70	71	72

P24	40mg/ml				
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	66
67	68	69	70	71	72

III.5. SDS ladders

HMW-SDS ladder (Amersham biosciences)



Appendix IV: Papers

Current Medicinal Chemistry, 2005, *12*, 2413-2446 1

Structure/Function Overview of Proteins Involved in Iron Storage and Transport

Peter J. Sargent, Sebastien Farnaud and Robert W. Evans*

Metalloprotein Research Group, Randall Research Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, Guy's Campus, London SE1 1UL, UK

Abstract: Iron, the major trace element in the body, is an essential component of many proteins and enzymes. As low-molecular-weight iron is potentially toxic to cells, higher organisms express a number of proteins for the transport and storage of iron. We review our current understanding of the intestinal absorption of iron in the light of recently identified membrane proteins, namely the ferric reductase, Dcytb, the two iron(II) transport proteins, DMT1 and ferroportin/Ireg1, and hephaestin, the membrane-bound homologue of the ferroxidase ceruloplasmin. Two types of mammalian transferrin receptor, TfR1 and TfR12, are now known to exist. The structure of TfR1 and its role in the process of receptor-mediated cellular uptake of iron are presented together with structural information on the iron storage protein ferritin. Mechanisms for the regulation of levels of TfR1 and ferritin, as well as other proteins involved in iron homeostasis, are discussed. Our current knowledge and understanding of the structure of members of the transferrin family of iron-binding proteins and the nature of the iron-binding centres in transferrins is presented, together with information on the processes of iron-uptake and iron-release by transferrin and a summary of the elements that have been found to bind to transferrins.

Keywords: Transferrin, ferritin, transferrin receptor, iron metabolism.

INTRODUCTION

With rare exceptions, virtually all organisms from *Archaea* to man are dependent on iron for survival [1]. The functional roles for iron include provision of a specific binding site for oxygen in the haem moiety of haemoglobin and conferment of redox activity on the cytochromes of the respiratory chain and on numerous enzymes, some of which are involved in DNA synthesis [2]. However, even though iron is an essential trace metal it is also toxic under certain conditions, for example in its non-chelated, labile form. Cellular iron deficiency is known to arrest cell growth and ultimately lead to cell death [3]. Organisms have therefore evolved highly sophisticated mechanisms to control iron homeostasis in order to provide the sufficient iron to support life, whilst at the same time minimising its toxic effects. In mammals this regulation is at many levels.

The daily iron requirement mainly for haemoglobin synthesis in a healthy individual is approximately 25mg. This varies from person to person. For instance, growing children and menstruating women have much higher iron demands. This demand increases even more for pregnant individuals. Iron requirements are three times higher in pregnancy than they are in menstruating women [4]. On average though, the production of 200 billion new erythrocytes every day requires ~20mg of iron for haemoglobin synthesis, accounting for nearly 80% of the iron demand [3]. The remaining 20% of the iron is used as a constituent in the production of other haemoproteins, iron-sulphur (Fe-S) proteins, and proteins that use iron in other functional groups.

*Address correspondence to this author at the Metalloprotein Research Group, Randall Division of Cell and Molecular Biophysics, King's College London, Guy's Campus, London SE1 1UL; Tel: +44-20-7848-6481; Fax: +44-20-7848-6485; E-mail: robert.evans@kcl.ac.uk

0929-8673/05 $50.00+.00

A typical Western diet provides on average 6mg of this iron (both haem and non-haem) per 4120 kJ of energy intake [5]. However, the amount of iron absorbed is regulated depending on the iron status of the individual. An individual with high levels of iron will not absorb iron so readily as someone with low levels. Also, the form of iron is important. Food iron occurs in two forms, haem iron and non-haem iron. Nearly 50% of iron in meat is of the haem form [4]. This form is the most important dietary source of iron, as it is more readily absorbed than non-haem iron. Hence vegetarians can be at a relatively greater risk for iron deficiency [5].

In normal individuals total body iron amounts to approximately 35 and 45mg/kg of body weight in adult women and men, respectively [6]. The majority of the iron, 60-70%, is present in haemoglobin in circulating erythrocytes, 10% in the form of myoglobin, cytochromes, and iron-containing enzymes, and the remaining 20-30% is stored in ferritin and haemosiderin in hepatocytes and reticuloendothelial macrophages [7]. This distribution of iron varies, for example in situations of iron deficiency, erythroid precursors have priority and the production of red blood cells continues at the expense of other tissues. The body is, however, economical in its handling of iron. Only 0.5-2mg enters and leaves the body on a daily basis [2] thus the majority of the iron is recycled.

IRON ABSORPTION AND TRANSPORT

Intestinal absorption of iron is thought to be the critical step in iron homeostasis [8]. The enterocytes found in the epithelial cell layer of the duodenum are specialised for absorption and transportation of iron. These cells differ from other cells in that they express a number of proteins related to iron uptake and transport (Fig. 1).

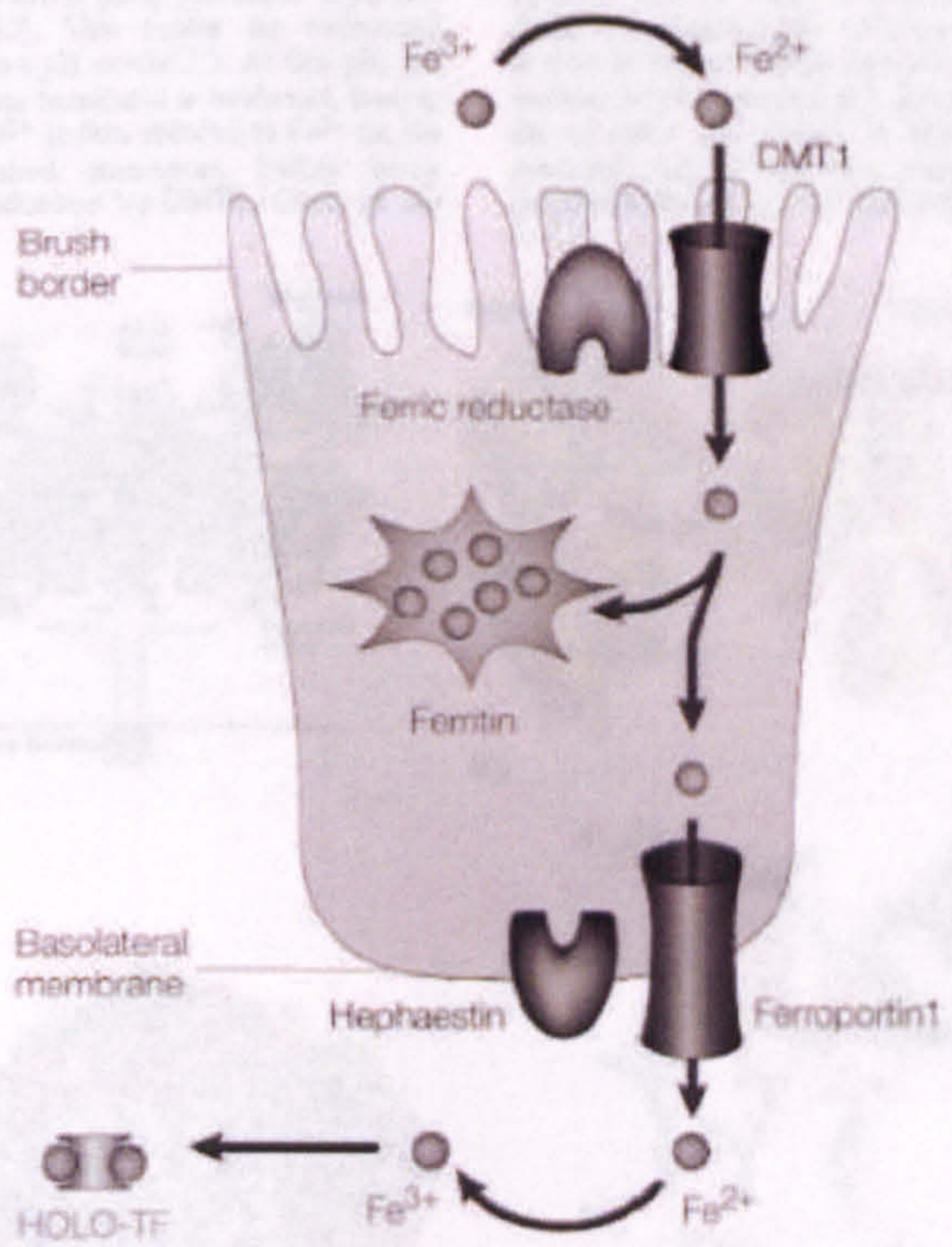


Fig. (1). A model for iron transport across duodenal epithelial cells
Proteins shown are those proposed to be involved in the absorption and transport of iron from the gastrointestinal tract (diet) into the circulation. Adapted from [2].

It has been proposed that non-haem iron (Fe^{3+}) in the duodenum is reduced by a ferric reductase (Dcytb: duodenal cytochrome b) in the brush border, before being transported into the enterocyte through the transmembrane iron transporter DMT1 (divalent metal transporter 1). Once inside the cell, the iron (Fe^{2+}) is either stored in ferritin, or passes through the basolateral membrane to reach the plasma by an iron exporter known as ferroportin or Ireg1. This step is thought to be followed by the reoxidation of Fe^{2+} by the plasma ferroxidase ceruloplasmin (or its membrane-associated homologue, hephaestin) [9]. Fe^{3+} is then sequestered up by apo-transferrin, although the details of this step are not yet fully understood. In normal individuals, this extracellular iron circulates in the plasma bound to serum transferrin (Tf) and in this form is no longer toxic, as it is unable to generate free radicals.

CELLULAR IRON UPTAKE

In vertebrates, the major protein involved in iron uptake of iron-loaded Tf into cells is the plasma membrane transferrin receptor (TfR1) [8]. Two types of TfR1, TfR1 and TfR12, are known. The latter molecule has only recently been identified and its function remains unclear [10]. It is known, however, that both receptors bind diferric transferrin in a pH-dependent manner and deliver iron to the cells. TfR12 has a much lower binding affinity than TfR1, approximately 25 times lower, and its distribution around the body varies significantly [11]. The best characterised of the two is TfR1 [12], a 760-amino-acid glycoprotein. The functional receptor is composed of two identical monomers, linked by two disulphide bridges to form a molecule of molecular mass 190,000 Da. The crystal structure of the ectodomain of the TfR1, resolved at 3.2Å, [13] is shown in Fig. (**2A**). Virtually all cells, except mature red blood cells, have this receptor on their surface, but the largest numbers are in the erythron, placenta and liver [14].

At physiological pH, this transferrin receptor has a high binding affinity for diferric serum transferrin. Following binding of the two proteins, the complex internalises through clathrin-coated pits into specialised endosomes (Fig. 3). Upon maturation and loss of the clathrin coat, the

endosome becomes competent to pump protons in a process energised by ATPase [15]. This causes the endosomal lumen to rapidly acidify to a pH around 5.5. At this pH, the binding of the iron to serum transferrin is weakened, leading to iron release. The free Fe^{3+} is then reduced to Fe^{2+} on the *cis*-side of the endosomal membrane, before being transported out of the endosome by DMT1. Once in the cytosol, iron is either stored or utilised by the cell. Following release of the iron, the transferrin-TfR1 complex is then recruited through exocytic vesicles back to the cell surface. At physiological pH, apo-transferrin dissociates from the receptor and returns to the circulation where it is reutilised [16]. It has been calculated that transferrin is recycled between 100-200 times during its lifetime [1].

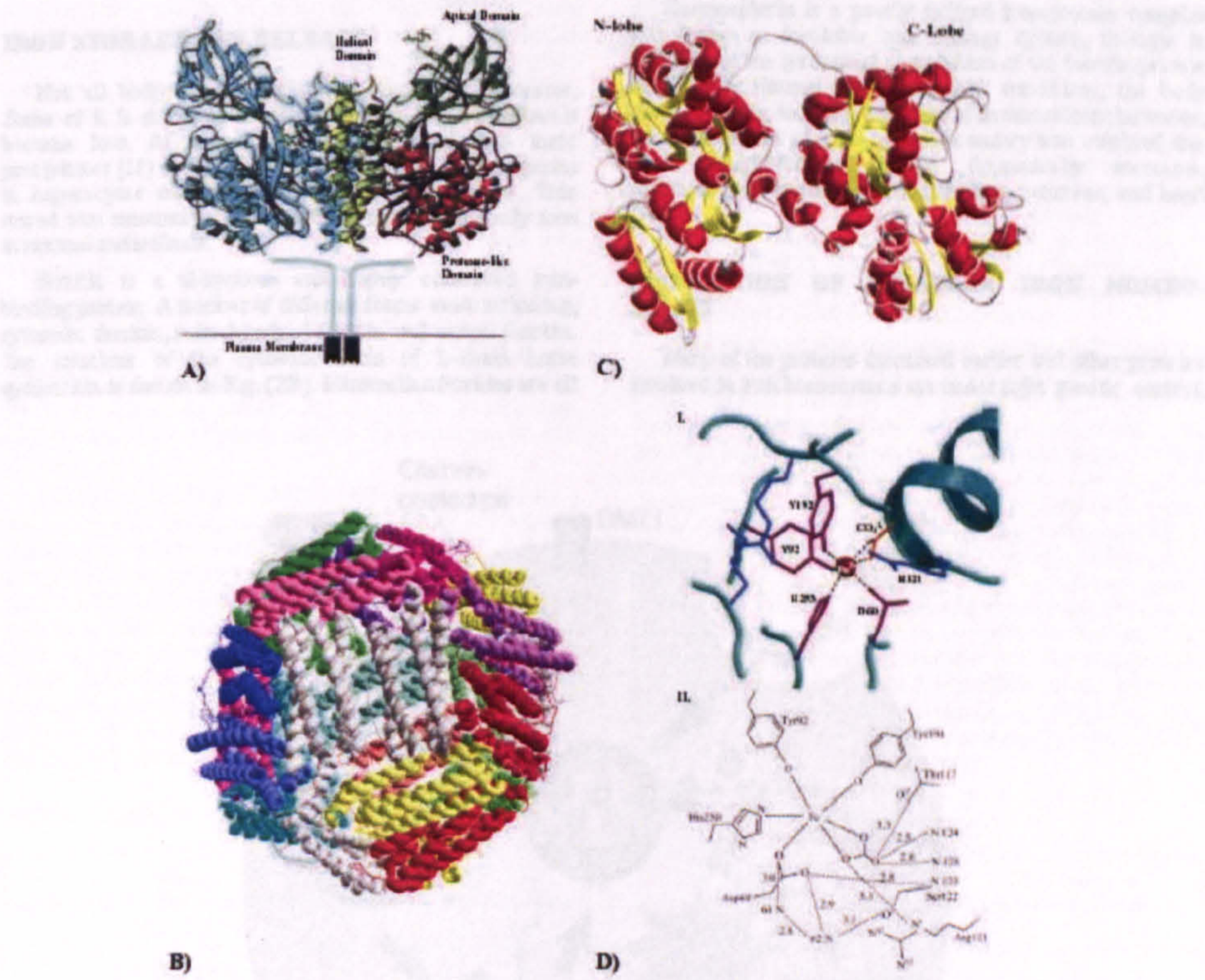


Fig. (2). **A)** The crystal structure of the ectodomain of the transferrin receptor (TfR1) resolved at 3.2 Å

Each TfR1 monomer can be divided up into a number of domains; Cytoplasmic domain, N-terminus consisting of 61 amino acids; Transmembrane domain, 26 amino acids long hydrophobic region; and the transferrin binding domain, 671 amino acids long. One monomer is depicted in blue, whilst the other monomer is shown in red, green yellow depicting the protease domain, apical domain and helical domain respectively. Adapted from [13].

B) A ribbon diagram of L-chain horse apoferritin (1.95Å) (1AEW) created using Deep Viewer / Swiss PDB Viewer

Each of the 24 subunits is coloured differently to give an idea of the complexity of the molecule.

C) Ribbon diagram of rabbit serum transferrin (1JNFA)

Picture was created using Deep Viewer / Swiss PDB Viewer. Both lobes are labelled and both domains highlighted. Domain I: Blue and Domain II: Red.

D) I. The N-lobe binding-site of human lactoferrin and II. Schematic diagram of the co-ordination environment and the hydrogen bonding network around Fe^{3+} in the N-lobe of hen serum transferrin

The Fe^{3+} ion is bound to the side chains of Asp60, Tyr92, Tyr192, and His253 and the bidentate CO_3^{2-} ion, which is hydrogen-bonded to Arg121 and to peptide NH groups from the N-terminus of an α-helix.

Cells may also take up iron in other protein-bound forms, such as acidic isoferritin, or haemoglobin, or through phagocytosis of red blood cells by macrophages. Biochemical data also suggests that additional iron uptake mechanisms may exist [3]. This demonstrates that even though the transferrin-mediated endocytosis is thought to be the major uptake system of iron, cells have evolved a number of other mechanisms for the uptake of iron.

IRON STORAGE AND RELEASE

Not all body iron is utilised in metabolic processes. Some of it is stored as a reserve in case body iron levels become low. As free iron would aggregate into toxic precipitates [17] iron is stored in ferritins and haemosiderins in hepatocytes and reticuloendothelial macrophages. This stored iron amounts to about 20-30% of the total body iron in normal individuals.

Ferritin is a ubiquitous and highly conserved iron-binding protein. A number of different forms exist including, cytosolic ferritin, mitochondrial ferritin and serum ferritin. The structure of the cytosolic form of L-chain horse apoferritin is shown in Fig. (**2B**). Mammalian ferritins are all heteropolymers of 24 subunits of two types, designated H for heavy (21,000 Da) or heart, and L for light (19,500 Da) or liver [1]. These molecules have the ability to sequester up to approximately 4500 iron atoms in the form of Fe^{3+} [18]. An important feature of ferritin is its ability to catalyse the oxidation of Fe^{2+}, the predominant iron form in the cytosol, converting it to Fe^{3+} before it is sequestered in the ferritin mineral core.

Haemosiderin is a poorly defined iron-protein complex that forms an insoluble iron storage system, thought to derive from the lysosomal degradation of the ferritin protein shell. Under normal iron homeostatic conditions, the body tissues contain only low amounts of haemosiderin; however, under conditions of primary and secondary iron overload, the tissue haemosiderin content dramatically increases, particularly in organs such as the liver, pancreas, and heart [19].

REGULATION OF CELLULAR IRON HOMEOSTASIS

Many of the proteins discussed earlier and other proteins involved in iron homeostasis are under tight genetic control.

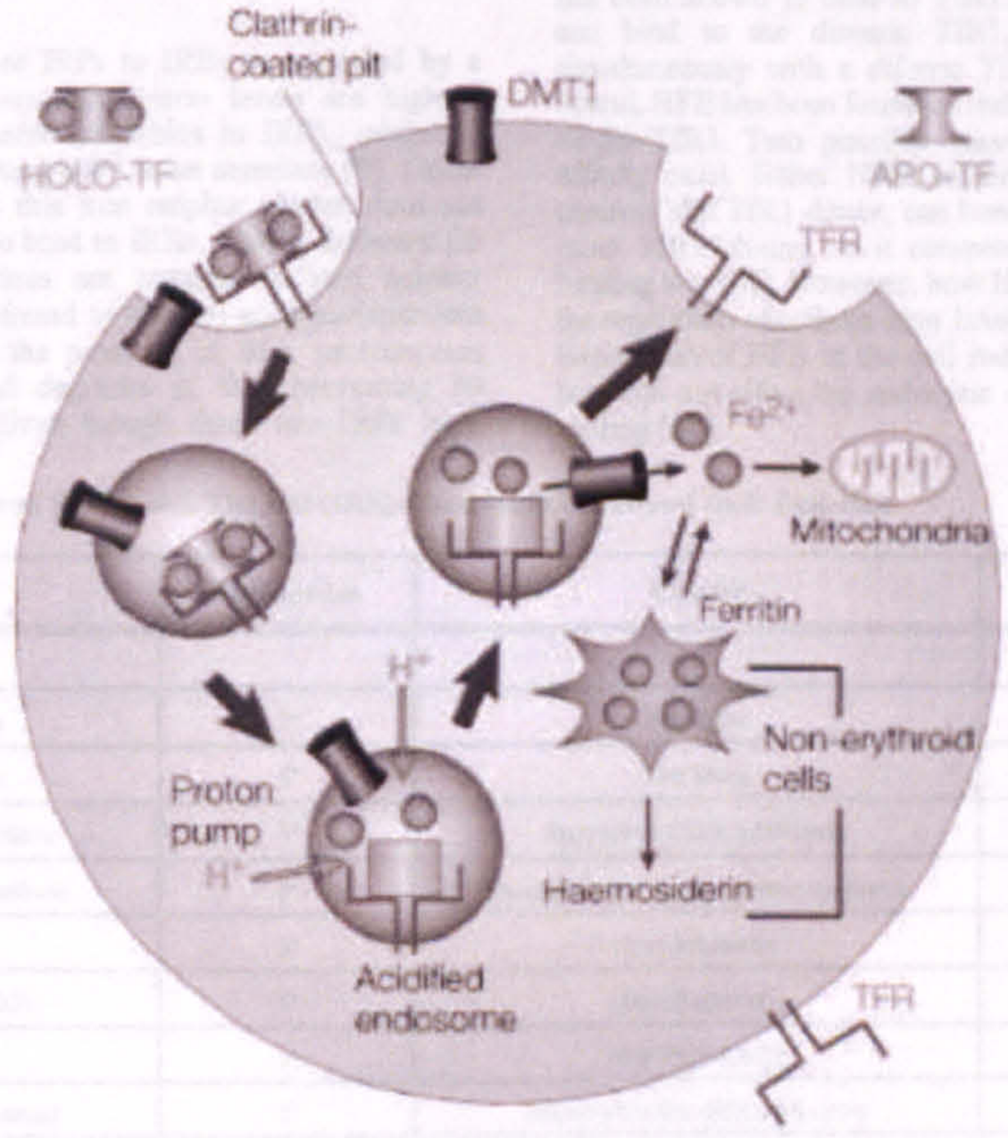


Fig. (3). A model for iron uptake via the transferrin receptor-mediated endocytosis pathway
Diferric transferrin is internalised upon binding to TFR1. Upon its release from transferrin, iron is transported into the cytosol for either utilisation or storage in ferritin. The transferrin-TFR1 complex is returned to the cell membrane transferrin is released back into the circulation. Adapted from [2].

Their expression can be up- or down-regulated depending on the availability of iron. This control involved modulation of transcription, mRNA stability, translation, and post-translational modification.

Post-transcriptional modification is the best characterised. This involves the iron regulatory proteins (IRPs) and iron responsive elements (IREs). IRP1 and IRP2 are involved in the coordinate posttranscriptional regulation of iron metabolism by binding to IREs [20] These IREs are stem-shaped structures located in the 5' and 3' untranslated regions (UTRs) of mRNA encoding proteins. Their structure includes a terminal hexanucleotide loop region with the sequence 5'-CAGUGX-3', where X can be A, C, or U, but never G, and a base-paired stem structure that is interrupted by a conserved unpaired cytosine, six nucleotides down from the loop. Table 1 shows examples of some of the proteins under IRE control. These include proteins involved in iron storage, erythroid haem synthesis, the TCA cycle, iron export and iron uptake.

The binding of IRPs to single IREs in the 5' UTRs of mRNA has been found to block translation. Whereas, binding of IRPs to IREs in the 3' UTRs stabilises the mRNA, therefore having a positive effect on translation. Fig. (4) shows some of the proteins that are regulated in this way, and also the effect iron levels have on the translation of their mRNAs.

The actual binding of IRPs to IREs is regulated by a number of different factors. When iron levels are high, a cubane iron sulphur cluster assembles in IRP1, inhibiting IRE binding and converting IRP1 to an aconitase [9]. Under iron limiting conditions this iron sulphur cluster does not form, and IRP1 is able to bind to IREs. This is different for IRP2 as this protein does not contain an iron sulphur cluster. IRP2 has been found to contain an iron-dependent degradation domain. In the presence of iron, proteosomes bind to this protein and degrades it, thus preventing its interaction with IREs. Even though these two IRPs have been identified, little is known about their individual contribution to the maintenance of cellular iron homeostasis. It has been shown *in vitro* that both proteins do not differ in their ability to be able to regulate ferritin and TfR1 expression [24]. However, it has been suggested [25] that IRP2 dominates post-transcriptional regulation of iron metabolism in mammals. The actual control of IRP expression is poorly understood. A number of factors that influence their expression are known. These include the influence of reactive oxygen species, shown to disassemble the iron sulphur cluster of IRP1, as well as nitric oxide and hypoxia having an effect on both IRPs [3].

Many other regulatory mechanisms exist that are independent of the IRP/IRE mechanisms. A number of cytokines have been reported to have an effect on the expression of some of the proteins involved in iron homeostasis. For example, tumor necrosis factor-α, interleukin-1, interleukin-6, and interferon-γ all stimulate H-ferritin expression but reduce TfR1 expression [26]. Interferon-γ has also been shown to induce DMT-1 expression and inhibit ferroportin expression.

Another example of regulation is the proposed role of the haemochromatosis protein HFE in iron homeostasis. HFE is a heterodimeric membrane protein that has the ability to form noncovalent associations with β_2-microglobulin and has been shown to bind to TfR1. Two molecules of HFE can bind to the dimeric TfR1, or just one molecule simultaneously with a diferric Tf. When both proteins are bound, HFE has been found to reduce the binding affinity of Tf for TfR1. Two possible reasons for this reduction in affinity exist. Either HFE, on binding to one polypeptide chain of the TfR1 dimer, can transmit structural changes to other TfR1 chains, or it competes with Tf for the TfR1 binding site [27]. However, how HFE actually contributes to the regulation of cellular iron homeostasis is still unknown. Expression of HFE in the cell reduces iron uptake by 33%, but does not affect the endocytic or exocytic rates of TfR1 cycling [28].

Table 1. Examples of Iron Responsive Element (IRE)-Containing Genes and their Function

Gene	IRE location	Function	Reference
Single IREs			
Ferritin L chain	5'	Iron Store	[21]
Ferritin H chain	5'	Iron Store	
Mitochondrial Aconitase	5'	Enzyme in Citric Acid cycle	
Erythroid 5-ALA-synthase	5'	Specific key enzyme in heme synthesis	
Nramp2	3'	Iron Importer	[22]
Ferroportin 1 (IREG1)	5'	Iron Exporter	[23]
DMT1	3'	Iron Transporter	[3]
Succinate dehydrogenase	5'	Involved in the citric acid cycle	[9]
Multiple IREs			
TfR1	3'	Trf-binding and –transport	[21]

The location of the IRE, whether in the 3' or 5' untranslated region (UTR), is also stated. Only the TfR1 mRNA contains multiple IREs.

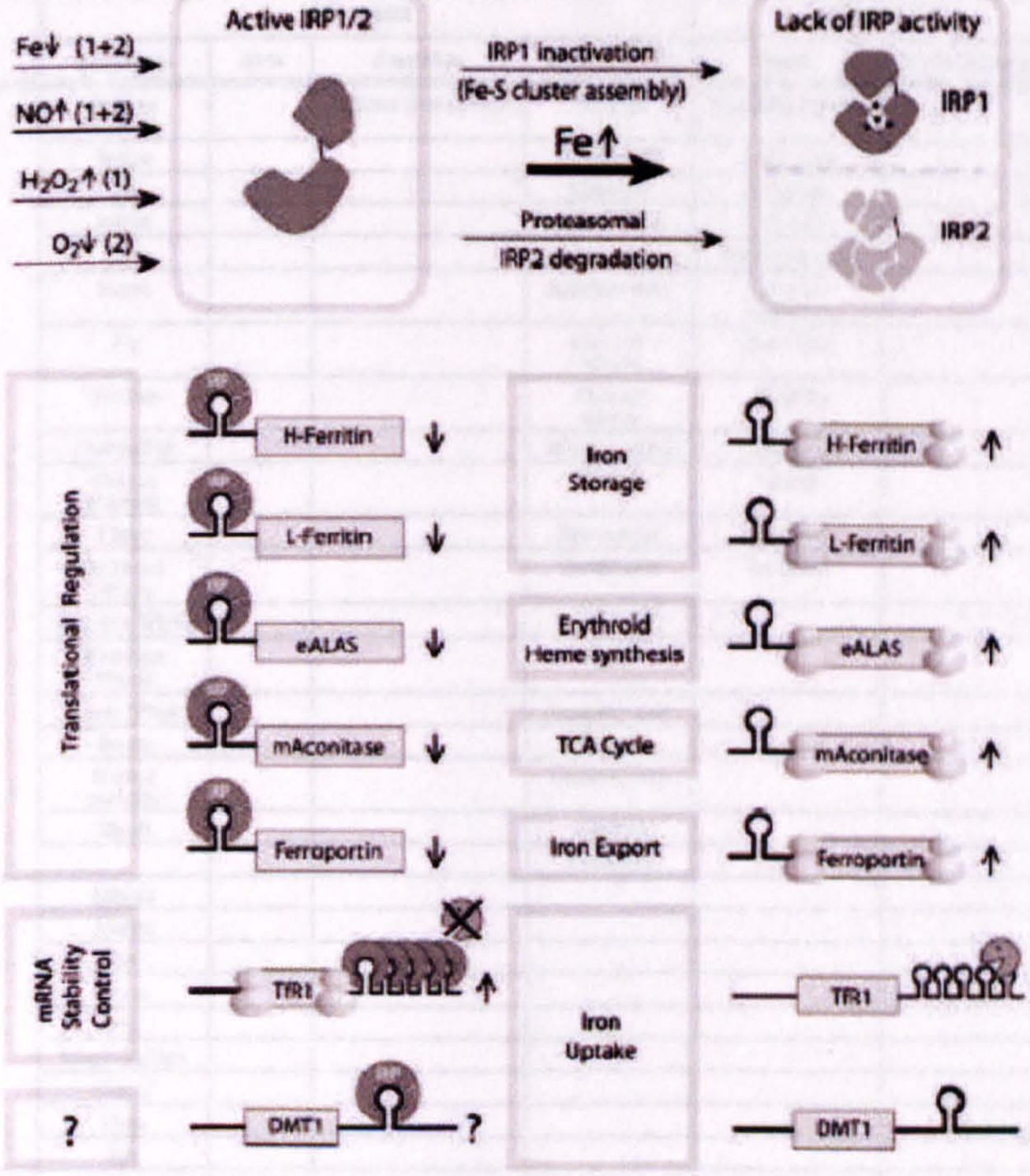


Fig. (4). The IRP-IRE regulation system showing the effect iron, NO, H_20_2 and O_2 levels has on the binding of IRPs to IREs, and the subsequent effect this has on translation on a number of proteins

IRPs binding to IREs in the 3' UTR portion of mRNA has been found to protect the mRNA from degradation, whereas, the binding of IRPs to IREs in the 5' UTR portion has been found to block translation. Adapted from [3].

TRANSFERRINS

The transferrins are a family of homologous monomeric bilobal glycoproteins of molecular mass approximately 80kDa that have an important role in iron homeostasis. Even though the first isolation of a member of the transferrin family, ovotransferrin, came in 1900 [29] it was not until 1944 that it a role was proposed for it [30]. Soon after, a similar protein, now known as serum transferrin, was isolated from blood plasma [31]. The transferrin family comprising four main members, each exhibiting a high affinity for iron (K_d~ 10^{-22}M) [32], includes the serum transferrins, which transport iron from its site of absorption to the site of utilisation or storage, the lactoferrins, which possess antimicrobial properties, the ovotransferrins, a major component of egg white, and the melanotransferrins, found on the surface of melanoma cells. These proteins are found in a wide range of organisms, in both vertebrate and invertebrate groups. Table **2** lists some of the transferrins that have either fully or partially sequenced.

General Structure

The transferrin polypeptide is approximately 670 amino acids long folded into two lobes that exhibit high levels of homology, the N- and C-lobe. The two lobes of human serum transferrin are 48% homologous [33] and for both porcine and rabbit serum transferrin, 45.5% and 42.9% respectively [34]. This high degree of homology is consistent with the theory that both lobes arose from a gene duplication process [35].

Table 2. Transferrins that have been Identified and Sequenced

Transferrin	Chordata				Anthropoda		Echinodermata
	Mammalia	Aves	Amphibia	Actinopterygii	Insecta	Malacostraca	Echinoidea
Serum Transferrin	Human		African clawed toad	Medaka	Mosquito (Aedes)		
	Mouse			Atlantic cod	Cockroach		
	Rat			Haddock	Silkworm		
	Rabbit			Coho salmon	Fruit fly		
	Cow			Cherry salmon	Picture wing fly		
	Horse			Rainbow trout	Tobacco hornworm		
	Pig			Sockeye salmon	Bean bug		
	Possum			Chinook salmon	Flesh fly		
	Guinea Pig			Atlantic salmon	Honeybee		
	Golden Hamster			Halibut	Termite		
	Hippo			Brown trout	Wax moth		
	Bowhead Whale			Brook trout	Tse tse fly		
	Cuvier's Whale			Lake trout			
	Peruvian Whale			Amago			
	Sperm Whale			Japanese char			
	Boutu			Gila trout			
	Harbor porpoise			Common carp			
	Sheep			Goldfish			
				Zebra fish			
Lactoferrin	Human						
	Mouse						
	Cow						
	Horse						
	Pig						
	Water buffalo						
	Camel						
	Goat						
	Rat						
Ovo-transferrin		Duck					
		Chicken					
Melano-transferrin	Human						
	Mouse						
	Chicken						
	Rabbit						
				Rainbow trout			
				Fugu			
Others							Red urchin
							Green urchin
							Sea urchin
							Hawaiin urchin
			Saxiphilin (bullfrog)				
						Pacifastin (crayfish)	

Partial and variants are shown in italics. This data was last updated in 2003 from: www.chatham.edu/undergraduate/bio/lambert/transferrin/transferrins.htm

Within the secondary structure of these proteins, there are six conserved disulphide bridges in each lobe, with an extra three semi-conserved bridges in the C-lobe [35]. Chicken and duck ovotransferrin have only fifteen bridges, whereas human lactoferrin has a further disulphide bridge in the C-lobe and human Tf has a further four bridges, two in the N-lobe and two in the C-lobe.

Since the first report of the structure of transferrins in 1979, from a 6Å electron density map [36], the tertiary structure of these proteins has been elucidated at much higher resolution. For example, the crystal structure of holo hen ovotransferrin N-lobe has been elucidated recently at 1.65Å [37], human apo-lactoferrin at 2.0Å [38] and rabbit serum transferrin at 2.6Å [31]. These structures show that the transferrins are folded into two lobes of approximately 330 amino acids that are joined together by a short connecting peptide and also make contact via a 'cushion' of hydrophobic residues packed between them. Each lobe can be further split into two dissimilar domains (CI, CII, NI and NII) (Fig. 2C). These domains consist of mixed right-handed β-sheets (~25%) and α-helices (~40%).

Most of the transferrins, except for *Xenopus laevis* transferrin, contain a varying number of N-linked glycosylation sites. For example, rabbit serum transferrin has only one [39], bovine lactoferrin has four [40] and human serum transferrin has two [41]. No O-linked glycosylations have been identified on transferrins. The fact that not all transferrins are glycosylated suggests that there is no specific role for these sugars.

Table 3. Comparison of the Residues Involved in Iron Binding, their Position and the Amino Acid Either Side of It (-1 and +1) for Both Lobes Amongst Different Transferrins

Transferrin	C-Lobe			N-Lobe		
	-1	Fe^{3+} ligand	+1	-1	Fe^{3+} ligand	+1
Human Serum Transferrin	Leu	Asp63	Ala	Leu	Asp392	Gly
	Phe	Tyr95	Tyr	Gly	Tyr426	Phe
	Gly	Tyr188	Ser	Gly	Tyr517	Thr
	Ser	His249	Thr	Asn	His585	Ala
	Gly	Arg124	Ser	Gly	Arg456	Thr
Human Lactoferrin	Leu	Asp60	Gly	Leu	Asp395	Gly
	His	Tyr92	Tyr	Gly	Tyr435	Leu
	Ser	Tyr192	Ser	Gly	Tyr528	Thr
	Ser	Tyr253	Ala	Asn	His597	Ala
	Arg	Arg121	Thr	Asp	Arg465	Thr
Rabbit Serum Transferrin	Leu	Asp63	Ala	Leu	Asp392	Gly
	Phe	Tyr95	Tyr	Gly	Tyr425	Leu
	Gly	Tyr188	Ser	Gly	Tyr514	Thr
	Ser	Tyr249	Ala	Asn	His582	Ala
	Gly	Arg124	Ser	Asp	Arg455	Thr
Porcine Serum Transferrin	Leu	Asp62	Ala	Leu	Asp396	Gly
	His	Tyr94	Tyr	Gly	Tyr431	Leu
	Gly	Tyr192	Ala	Gly	Tyr526	Thr
	Ser	His253	Ala	Asn	His594	Ala
	Gly	Arg123	Ser	Asp	Arg462	Thr
Camel Lactoferrin	Leu	Asp60	Gly	Leu	Asp395	Gly
	His	Tyr92	Try	Gly	Tyr433	Leu
	Gly	Tyr192	Ser	Gly	Tyr526	Thr
	Ser	His253	Ala	Asn	His595	Ala
	Gly	Arg121	Ser	Asp	Arg463	Thr

Residues highlighted in bold are those directly involved in the binding of the ferric ion. Out of these residues, italicised text represents the residues involved in carbonate binding.

Iron-Binding Sites

The transferrins have the ability to bind reversibly two ferric ions, one in each lobe. The two iron-binding sites are 42Å apart and buried deep in the interdomain cleft in a highly polar environment [42]. The iron-binding sites in both lobes are very similar and comprise a distorted octahedral coordination involving four protein ligands and a bidentate synergistic carbonate anion [34] (Fig. 2DI). These protein ligands comprise two tyrosine residues, one aspartate residue and one histidine residue. Amongst all known transferrins the aspartate and the histidine are provided by domain I, whereas the two tyrosines and the bidentate carbonate anion is provided by domain II. All four of these ligands, along with an arginine involved in the anchoring of the carbonate anion to the protein, are conserved between lobes and amongst higher transferrins. Table 3 illustrates to what degree these five residues, along with their neighbours, are conserved amongst the higher transferrins.

In order to understand the binding of Fe^{3+} by transferrin, it is important to note that Fe^{3+} has the ability to bind six ligands. This arises from the fact that the atom has five unpaired electrons in its 3d orbital and an empty 4s orbital. When Fe^{3+} is bound by transferrin, a possible way these two orbitals are filled up is by an electron from one carboxylate oxygen (aspartate), an electron from each of the two phenolate oxygens (two tyrosines), an electron from each of the two oxygens from the carbonate anion, and two electrons from the imidazole nitrogen (histidine) (see Fig 2DII) It is also important to note that the three positive charges of the Fe^{3+} are matched by the 3 negative charges of the aspartate and the two tyrosines. Also, the charge on the anion is compensated for by the positively charged arginine side chain and a positive charge at the N-terminus of an α-helix [43].

Iron Binding and Release

A feature common to all transferrins is the closing of the interdomain clefts upon incorporation of iron into the iron-binding sites, bringing the iron-binding ligands on both domains together, 'grasping' the iron. In the apo form, the domains have been shown to be open [44].

The actual mechanism of iron binding is still not completely clear. It is thought that iron-binding starts with the binding of the carbonate anion to domain II, followed closely by the binding of the metal by the two tyrosines on the same domain. Once bound, a conformational change occurs, causing domain I to close over domain II This closure allows the histidine and aspartate ligands to come into contact with the iron, allowing them to bind the irons remaining two coordinates.

The extent in which the two domains open and close is reported in a recent study [45] which also demonstrates how highly conserved the rotation angle of domains is across the transferrin family When domain I from the closed form of hen serum transferrin N-lobe is superimposed with the open form, a rotation of 52.7° difference can be seen. Similarly in the C-lobe, a rotation of 35° was seen. The rotations in the N-lobe of human lactoferrin and human serum transferrin are also similar, 54° and 63° respectively.

The actual affinity for the Fe^{3+} ion varies between both lobes and also members of the transferrin family. One explanation for this is that the iron-ligand bond distances are known to vary. The longer bond length is thought to be the reason why the N-lobe of serum transferrin releases iron more readily than the C lobe on increasing acidity. Iron is released from the N-lobe of serum transferrin at around pH 5.7, whereas the C lobe retains iron until around pH 4.8 [46]. A difference in the pH dependence of iron release can also be seen between transferrin family members as, for example, lactoferrin retains iron at a pH as low as 3.5. The higher iron-binding affinity seen in lactoferrins could also be due to the greater number of interdomain hydrogen bonds [34].

These differences in iron affinity between the serum transferrins and the lactoferrins has also been proposed to be due to particular residues found only in the N-terminal lobe of the serum transferrins, away from the iron-binding site, at the interdomain interface. In this region, two lysines have been proposed to play a role in iron-release. These residues are on opposite faces of the iron-binding cleft and are thought to be bridged by a single proton when the protein is in its closed form. The theory is that at low pH these two lysines are protonated, leading to the subsequent release of the proton bridge and causing the two domains to repel each other [4] As the domains flip open, the iron is released. This theory is known as the dilysine trigger. The importance of these lysines was highlighted in a study [47] that showed that mutation of these two lysines had a significant effect on iron binding and release.

However, a primary role of this dilysine trigger has been questioned. It is thought that the protonation of the carbonate anion and the histidine iron-binding ligand constitutes the primary stages of iron release, with the dilysine trigger being important once the iron binding ligands are weakened [34]. This dilysine trigger does explain though, at least in part, the *in vitro* release of iron more readily from the N-lobe compared to the C-lobe of serum transferrin.

Binding of Other Metals

Although the transferrins have evolved to bind iron (III) with high affinity at neutral pH, their metal-binding sites are able to bind a wide range of different metal ions. As serum transferrin in normal individuals is only partially iron-saturated transferrin could in theory bind and transport metals other than iron that could ultimately be taken up into cells by receptor-mediated endocytosis.

By 1992 thirty elements, in addition to iron, were known to be able to bind to one or both of the metal-binding sites in human serum transferrin *in vitro* [48]. Since then a further four metal ions, cerium (IV) [49], bismuth (III) [50], lutetium (III) [51] and ytterbium (III) [52] have been shown to bind to members of the transferrin family of proteins. The elements now proven to bind to transferrins are highlighted in Fig. (5). Despite the large number of metal ions that are now known to bind to the transferrins, crystal structures of transferrins complexed to metals other than iron (III) are limited to human lactoferrin with copper (II) [53] and cerium (IV) [54], and mare lactoferrin with samarium (III) [55]. The availability of further crystal structures for metal complexes

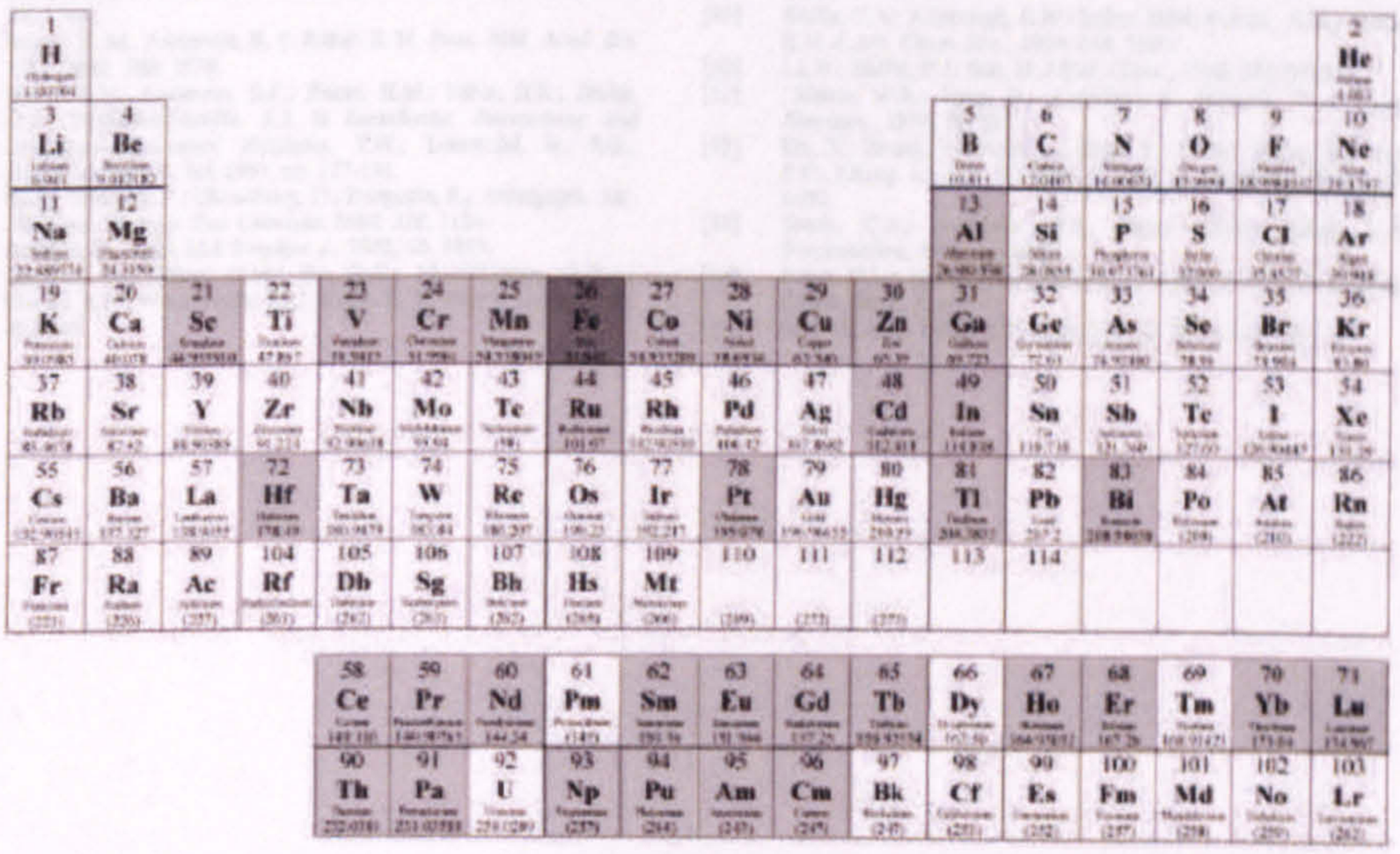


Fig. (5). The Periodic Table showing the elements (shaded) that binds to one or both of the metal-binding sites of transferrins.

of transferrins would provide invaluable information on how this family of proteins is able to accommodate such a wide range of divalent, trivalent and tetravalent metal ions with different ionic radii.

ACKNOWLEDGEMENT

The authors thank the Medical Research Council of the United Kingdom and the Wellcome Trust for financial support.

REFERENCES

[1] Aisen, P; Enns, C.; Wessling-Resnick, M. *Int. J. Biochem. Cell Biol.*, 2001, *33*, 940.
[2] Andrews, N.C. *Nat. Rev. Genet.*, 2000, *1*, 208.
[3] Hentze, M.W; Muckenthaler, M.U; Andrews, N.C. *Cell*, 2004 *117*, 285.
[4] Tapiero, H; Gate, L; Tew, K.D. *Biomed. Pharmacother.*, 2001, *55*, 324.
[5] Beard, J; Tobin, B. *Am. J. Clin. Nutr.*, 2000, *72*, 394S.
[6] Liao, P.I; Hallahan, M.; Peterson, P.A.; Yang, Y. *Molecular Aspects Med.*, 2001, *22*, 1.
[7] Conrad, M.E; Umbreit, J.N; Moore, E.G. *Am. J. Med. Sci.*, 1999, *318*, 213.
[8] De Freitas, J.M; Munaghini, R. *Biomed. Res.*, 2001, *473*, 153.
[9] Pantopoulos, K. *Ann. N. Y. Acad. Sci. USA*, 2004, 1012, 1.
[10] Kawabata, H; Yang, R.; Hirama, T; Vuong, P.T.; Kawano, S; Gombart, A.F; Koeffler, H.P. *J. Biol. Chem.*, 1999, *274*, 20826.
[11] Trinder, D; Baker, E. *Int. J. Biochem. Cell Biol.*, 2003, *35*, 292.
[12] Cheng, Y; Zak, O; Aisen, P; Harrison, S.C; Walz, T. *Cell*, 2004, *116*, 565.
[13] Lawrence, C.M; Ray, S; Babyonyshev, M; Galluser, R; Borhani, D.W; Harrison, S.C. *Science*, 1999, *286*, 779.
[14] Baguin, Y. *Clin. Chim. Acta*, 2003, *329*, 9.
[15] Qian, Z.M; Li, H; Sun, H; Ho, K. *Pharmacol. Rev.*, 2002, *54*, 561.
[16] Morgan, E.H. *J. Gastroenterol. Hepatol.* 1996, 11, 1027.
[17] Aruoma, P; Laval, S. *Free Radic. Biol. Med.*, 2002, *33*, 457.
[18] Chasteen, N.D; Harrison, P.M. *J. Struct. Biol.*, 1999, *126*, 182.
[19] Theil, E. *Proteins of Iron Metabolism*. CRC Press, 2002.
[20] Wang, J; Pantopoulos, K. *Mol. Cell Biol.*, 2002, *22*, 4438.
[21] Hentze, M.W; Kuhn, L.C. *Proc. Natl. Acad. Sci. USA*, 1996, *93*, 8175.
[22] Wardrop, S.L; Richardson, D.R. *Eur. J. Biochem.*, 1999, *263*, 41.
[23] Mok, H; Jahnel, J.; Pai, S; Cathanach, B.M; Pinhal, J.I; Youssoufian, H; Schumacher, A. *Development*, 2004, *131*, 1839.
[24] Ishitsi, R; Long, J.C; Theil, E.C. *J. Biol. Chem.*, 2002, *277*, 42379.
[25] Mayron-Holtz, E.G; Ghosh, M.C; Iwai, K; LaVaute, T; Brazzolotto, X; Berger, U.V; Land, W; Ollivierre-Wilson, H; Grinberg, A; Love, P; Rouault, T.A. *EMBO J.*, 2004, *23*, 386.
[26] Torti, F.M; Torti, S.V. *Blood*, 2002, *99*, 3505.
[27] Giannetti, A.M; Björkman, P.J. *J. Biol. Chem.*, 2004, *279*, 25866.
[28] Roy, C.N; Penny, D.M; Feder, J.N; Enns, C.A. *J. Biol. Chem.*, 1999, *274*, 9022.
[29] Osborne, T.B; Campbell, G.F. *J. Am. Chem. Soc.*, 1900, *22*, 422.
[30] Schade, A.L; Caroline, L. *Science*, 1944, *100*, 14.
[31] Schade, A.L; Caroline, L. *Science*, 1946, *104*, 340.
[32] Aisen, P; Leibman, A; Zweier, J. *J. Biol. Chem.*, 1978, *248*, 649.
[33] Baldwin, G.S. *Comp. Biochem. Physiol. B*, 1993, *106*, 203.
[34] Hall, D.R; Haddan, J.M; Leonard, G.A; Bailey, S; Neu, M; Winn, M; Lindley, P.F. *Acta Cryst. D*, 2001, *57*, 70.
[35] Williams, J. *Trends Biochem. Sci.*, 1982, *7*, 394.
[36] Gorinsky, B; Horsburgh, C; Lindley, P.F; Moss, D; Pasher, M; Watson, J.L. *Nature*, 1977, *281*, 157.
[37] Mizutani, K; Mikami, B; Hirose, M. *J. Mol. Biol.*, 2001, *309*, 937.
[38] Jameson, G. B; Anderson, B.F; Norris, G.E; Thomas, D.H; Baker, E.N. *Acta Cryst. D*, 1998, *54*, 1319.
[39] Evans, R.W; Aitken, A; Patel, K.J. *FEBS Lett.*, 1988, *238*, 39.
[40] Baker, E.N; Lindley, P.F. *J. Inorg. Biochem.* 1992, *47*, 47.
[41] Jamieson, G.A; Jett, M; DeBernardo, S.L. *J. Biol. Chem.*, 1971, *246*, 3686.

[42] Anderson, B.F.; Baker, H.M.; Dodson, E.J.; Norris, G.B.; Rumball, S.V.; Waters, J.M.; Baker, E.N. *Proc. Natl. Acad. Sci. USA*, 1987, *84*, 1769.
[43] Baker, H M., Anderson, B. F. Baker. E. N. *Proc. Natl. Acad. Sci. USA*, 2003, *100*, 3579.
[44] Baker, E.N.; Anderson, B.F.; Baker, H.M.; Faber, H.R.; Smith, C.A.; Sutherland-Smith, A.J. In *Lactoferrin: Interactions and Biological Functions*. Hutchens, T.W.; Lonnerdal, B., Eds.; Humana: Totowa, NJ, 1997, pp. 177-191.
[45] Guha Thakurta, P.; Choudhury, D.; Dasgupta, R.; Dattagupta, J.K. *Biochem. Biophys. Res. Commun.* 2004, *316*, 1124.
[46] Rinaldo, D.; Field, M.J. *Biophys J.*, 2003, *85*, 3485.
[47] Nurizzo, D.; Baker, H.M.; He, Q-Y.; MacGillivray, R.T.A.; Mason, A.B., Woodworth,; R.C. Baker, E. N. *Biochemistry*, 2001, *40*,1616.
[48] Welch, S. *Transferrin: the Iron Carrier*. CRC Press: Boca Raton, 1992.
[49] Smith, C.A.; Ainscough, E.W.; Baker, H.M.; Brodie, A.M.; Baker, E.N. *J. Am. Chem. Soc.*, 1994, *116*, 7889.
[50] Li, H.; Sadler, P.J.; Sun, H. *J Biol. Chem.*, 1996, *271*, 9483
[51] Harris, W.R.; Yang, B.; Abdollahi, S.; Hamada, Y. *J. Inorg. Biochem.*, 1999, *76*, 231.
[52] Du, X.; Zhang, T.; Yuan, L.; Zhao, Y.; Li, R.; Wang, K.; Yan, S.C.; Zhang, L.; Sun, H.; Qian, Z. *Eur. J. Biochem.* 2002, *269*, 6082.
[53] Smith, C.A.; Anderson, B.F.; Baker, H.M.; Baker, E.N. *Biochemistry*, 1992, *31*, 4527.
[54] Baker, H.M.; Baker, C.J; Smith, C.A.; Baker, E.N. *J. Biol. Inorg. Chem.*, 2000, *5*, 692.
[55] Sharma, A.K.; Singh, T P. *Acta Cryst. D*, 1999, *55*, 1799.